

**A búzaszem fejlődésében szerepet játszó gének
azonosítása bioinformatikai és molekuláris
módszerekkel**

Doktori (Ph.D.) értekezés

SZÜCS ATTILA

**MTA Szegedi Biológiai Központ,
Növénybiológiai Intézet**

Témavezető: Dr. Fehér Attila

Szeged

2007

Tartalomjegyzék

Rövidítések és szakkifejezések jegyzéke	4
1. Bevezetés.....	6
2. Irodalmi áttekintés	8
2. 1. <i>A búzaszem fejlődése</i>	8
2. 1. 1. A búza virágzata és a szem felépítése	8
2. 1. 2. Megtermékenyítés és apomixis	9
2. 1. 3. A növényi megtermékenyítés molekuláris biológiája	10
2. 1. 4. A szemfeltöltődés folyamata	12
2. 2. <i>A szárazság és/vagy hőstressz hatása a növényekre</i>	14
2. 2. 1. A szárazság stressz	14
2. 2. 2. A magas hőmérséklet hatása a növényekre	16
2. 2. 3. A magas hőmérséklet és a szárazság kombinált hatása a növényekre	17
2. 3. <i>A szárazság és/vagy hőstressz hatása a búza ivaros folyamataira és a szemfejlődésre</i>	18
2. 3. 1. A pollen és az embriózsák kialakulása stressz körülmények között	18
2. 3. 2. Szemfeltöltődés hő és szárazság stressz alatt	19
2. 4. <i>Modern genetikai módszerek a növénynevelésben</i>	20
2. 4. 1. Molekuláris markerek	20
2. 4. 2. Funkcionális genomika	21
2. 4. 2. 1. A transzkripció profil meghatározására szolgáló technikák	21
2. 4. 2. 2. Genomikai megközelítések a gabonafélék stressztűrésének vizsgálatában	24
2. 4. 3. Modern génszűrés módszerek	25
3. Célkitűzések	27
4. Anyagok és módszerek	28
4. 1. <i>Programozás és bioinformatika</i>	28
4. 1. 1. A programok elkészítéséhez használt fejlesztői környezet	28
4. 1. 2. Az adatok beolvasása	29
4. 1. 3. Adatbázis műveletek	33
4. 2. <i>A klónkönyvtárak előállítása</i>	37
4. 3. <i>RT-PCR</i>	40
4. 3. 1. cDNS szintézis	40
4. 3. 2. Valós idejű kvantitatív reverz-transzkripció PCR (RT-QPCR)	40
5. Eredmények	45
5. 1. <i>Saját készítésű microarray kifejlesztése és tesztelése</i>	45
5. 1. 1. <i>A búza adatbázis felépítéséhez szükséges adatok</i>	45
5. 1. 2. Az érdekes gének kiválasztása és az oligonukleotid tervezés	48
5. 1. 3. A búza chip tesztelése	51
5. 2. 1. Agilent microarray chip tervezés hő, ill. hő és szárazság stressz a búzaszem fejlődésére gyakorolt együttes hatásának vizsgálatára	55
5. 2. 1. 1. Az oligonukleotidok tervezéshez és az eredmények kiértékeléséhez szükséges adatbázis elkészítése	55
5. 3. 1. Búza petesejt és korai (8 órás) zigóta transzkripció analízise	62
5. 3. 3. Pete és zigóta specifikus gének meghatározása RT-PCR technológiával	73

6. Eredmények megvitatása	77
6. 1. <i>A növényi petesejt transzkripció profilja</i>	77
6. 2. <i>A zigóta és a kétsejtes embrió transzkripció profilja</i>	78
6. 3. <i>Genomikai megközelítési lehetőségek a búza kutatásában</i>	79
6. 4. <i>Szekvencia adatok kezelése és feldolgozása</i>	82
7. Összefoglalás	83
8. Köszönetnyilvánítás	85
9. Hivatkozási lista	86
10. Ph. D. thesis	96
11. Publikációs lista	102
12. Függelék	103
12. 1. <i>Nagy homológiát mutató pete EST-k annotációja</i>	103
12. 2. <i>Nagy homológiát mutató zigóta EST-k annotációja</i>	108
12. 3. <i>Maghoz köthető EST könyvtárak listája</i>	121

Rövidítések és szakkifejezések jegyzéke

LEA	late embryogenesis abundant
ABA	abscisic acid
AGP	ADP-glucose pyrophosphorylase
BLAST	Basic Local Alignment Search Tool
cDNS	complementer DNS
DNS	dezoxiribonukleinsav
EC	enzyme catalogue
EST	expresses sequence tag
FRET	Förster resonance energy transfer
Gb	gigabyte
GBSS	granule-bound starch synthase
GI	gene identifier
GO	gene ontology
HSF	heat stress transcription factor
HSP	heat shock protein
IO	input-output
KEGG	Kyoto Encyclopedia of Genes and Genomes
Mb	megabyte
mRNS	hírvivő RNS
NCBI	National Center of Biotechnology
PCR	polymerase chain reaction
QTL	quantitative trait loci
RNS	ribonucleid acid
rRNS	riboszómális RNS
RT-PCR	Real-time PCR
RT-QPCR	Real-time quantitative PCR
SBE	starch branching enzyme
SQL	Structured Query Language

TA	Transcript Assemblies of ESTs
TC	Tentative Consensus sequence of contig
TIGR	The Institute for Genomic Research
UniRef	UniProt Reference Clusters
xml	Extensible Markup Language

1. Bevezetés

A búza (*Triticum* sp.) a világon a rizs és a kukorica mellett a legelterjedtebb haszonnövény. Míg a kukorica jelentőségét az állattenyésztés, és mostanában egyre jobban (várhatóan egyre növekvő ütemben) a bioetanol gyártás adja, addig a rizs és a búza főleg emberi táplálékként szolgál.

A rizs a trópusi, szubtrópusi országokban élő emberek fő energiaforrása, ezzel szemben a különböző búzafajok és fajták gyakorlatilag mindenütt termesztetők Fokvárostól Helsinki-ig. Természetesen a fajok, fajták termesztetőségi adatait jelentősen befolyásolják a helyi adottságok (nappalok hossza, napi hőmérsékleti ingadozás, átlagos hőmérséklet, csapadék viszonyok, talaj típusa stb.). A búzát Magyarországon évente kb. 1.000.000 hektáron vetik, átlagos hozama 4 tonna / hektár. A megtermelt búza értéke általában több mint 100 milliárd forint fölött van évente. A termelt búza több mint felét exportáljuk, ami a külkereskedelmi hiány csökkentésében is fontos.

Nem tudjuk, hogy az emberi tevékenység, vagy valamely természeti (pl. csillagászati) jelenség a fő oka, de tény, hogy az utóbbi évtizedekben jelentősen kezd megváltozni az éghajlat. Én, személy szerint, nem szeretem a „globális felmelegedés” kifejezést, mert nem tudatosítja az emberekben, hogy az éghajlatváltozás lokális szinten történik meg, ami azt is jelentheti, hogy míg Európa egyes területein nem változik meg jelentősen az éghajlat, máshol komoly változások történhetnek. Rendkívül nagy a bizonytalanság ebben a kérdésben. Nem lehet elegendően megbízható előrejelzéseket tenni az időjárás változására, de fel kell készülni rá. Cromwell szavaival élve “Put your trust in God; but be sure to keep your powder dry”, azaz “Bízzál istenben, de tartsd szárazon a lőport”.

A legvalószínűbb prognózisok szerint, Magyarország egyre inkább mediterrán klímájú ország lehet, ami azt jelenti, hogy a nyarak egyre szárazabbak és egyre melegebbek lesznek. Ezt a problémát felismerve, megkezdődtek hazánkban azok a kutatási programok, amelyek a szárazságtűrő búzafajták nemesítésének hatékonyabbá tételét célozzák meg, amely magában foglalja a búza genetikájának jobb megismerését, a különböző búzafajták stressz tűrő képességeinek vizsgálatát, a szárazság hatására lejátszódó biokémia folyamatok feltérképezését.

A Ph.D. értekezésemben azokról a genetikai kutatásainkról számolok be, amellyel mi is hozzájárulhatunk ahhoz, hogy megtaláljuk azokat a géneket, amelyek felelősek lehetnek a termés minőségért, mennyiségéért és nem utolsósorban a termésbiztonságért. A jelen dolgozatban vázolt megközelítéseket elsősorban a búzaszem fejlődés hő- és szárazságérzékenységének kutatására kívánjuk a közeljövőben felhasználni. Hosszú távon pedig abban reménykedünk, hogy alapkutatásaink eredményeire támaszkodva a jövőben lehetővé válhat a hosszadalmas búzanemesítési folyamat lerövidítése, a szelekció hatékonyságának növelése.

2. Irodalmi áttekintés

2. 1. A búzaszem fejlődése

2. 1. 1. A búza virágzata és a szem felépítése

A búza (*Triticum aestivum* L.) virágzata kalászvirágzat, a kalászkák a kalászsorsón vannak és két kalászkapelyva fogja közre, védi. A kalászka általában 3-5 virágból áll. A külső virágpelyva gyakran szálkában (toklász) végződik. A virágpelyvák közt van a termő és a három porzó. A kalászon redők jelzik a differenciálódó kalászkák helyét és számát. Leghamarabb a kalász közepe táján levő kalászkák differenciálódnak, majd az alapi és végül a csúcsi kalászkák alakulnak ki. Ez a sorrend figyelhető meg virágzáskor és éréskor is. A búza termése szemtermés, a terméshéj és a maghéj összenő. A három rétegből összenőtt terméshéj és a kétrétegű maghéj védi az embriót (csíra) és az endospermiumot (táplálósövet). A csíra a megtermékenyített petesejtből fejlődik, míg a megtermékenyített központi sejt magjából a triploid táplálósövet, a későbbi mag endospermiuma alakul. Az endospermium egy külső aleuron rétegből és belső, keményítőt tartalmazó sejtekből áll. A csíra és az endospermium közt helyezkedik el a scutellum, ami közvetíti a táplálósövetben felhalmozódott keményítőt, fehérjét, cukrot stb. a növekvő csírához. A szemben összegyűlt keményítő és a protein tartalom főleg az endospermium sejtjeinek számától függ, amely már a szemfejlődés korai szakaszában kialakulva alapvetően meghatározza a realizálható termés mennyiségét (Egli, 1998). Amíg a megtermékenyítés hatékonysága a szemszámot, addig a szemfeltöltődés a szem méretét és minőségi jellemzőit szabja meg. Ez utóbbi folyamat magába foglalja a szemben előforduló szénhidrátok, zsírsavak, fehérjék és ásványi anyagok szintézisét, transzportját és elraktározását. A keményítő a fő meghatározója a búza termés mennyiségnek, mivel mintegy 65-75 %-át adja a szem nedves tömegének és 80 %-át az endospermium száraz tömegének (Rahman és mtsai., 2000; Slattery és mtsai., 2000).

2. 1. 2. Megtermékenyítés és apomixis

A magfejlődésnek két alapvető fajtája létezik az ivaros és az ivartalan. Az ivaros megtermékenyítést követően meginduló magfejlődés, amely a kettős megtermékenyítéssel kezdődik, aminek eredményeként a petesejtből a zigóta, a központi sejtől a táplálósövet (endospermium) jön létre. Előbbiből az embrió fejlődik, míg utóbbi a csírázáshoz biztosítja az elraktározott keményítőt, fehérjét és lipideket.

Bizonyos esetekben azonban a megtermékenyítés hiányában is megindul a magfejlődés: ez a jelenség az apomixis. Az apomixis a természetben nem ritka, mintegy negyven különböző családba tartozó, több mint 400 faj esetében figyelték meg (Koltunow és mtsai., 1995), például a gyermekláncfű (*Taraxacum*-), a hölgymál (*Hieracium*-) fajokban, továbbá a havasi gyopár (*Leontopodium alpinum*), borsó (*Pisum sativum*) esetében is. Érdekességgént megjegyzem, hogy Mendel élete végén semmire se tartotta korszakalkotó genetikai kísérleteit, mert nem tudta megmagyarázni az apomixis jelenségét. Az apomixis egyik formája az aposporia, ahol az embrió nem a petesejtől, hanem a diploid embriózsák egyéb sejtjeiből, pl. az egyik segítősejtől (*Iris siberica*), vagy az antipodiális sejtek egyikéből (*Allium*-fajok) alakul ki. Ha az embrió a nucellus valamelyik diploid sejtjéből fejlődik sporofitikus apomixisről vagy járulékos embriógenézisről beszélünk (Koltunow, 1993).

Az apomixis gazdasági jelentőségét az adja, hogy míg az ivaros szaporodás genetikailag változatos utódokat eredményez, amelyek biztosítják a megfelelő változatosságot a faj túléléséhez és a nemesítéshez, azaz a természetes, valamint a mesterséges szelekcióhoz, addig ez a változatosság a termesztett növény fajták esetében már nem kívánatos (eltérő érésidő, más magasság, ami a betakarítást megnehezítheti). Termesztett növényeink egy részénél a leghatékonyabb termelési mód a hibrid fajták termesztésén alapul. Ilyenkor a vetőmag előállító cégek két beltenyésztett vonalat tartanak fent (nem olcsón), és a vetőmag előállítás során ezeket keresztezik. Amennyiben a vonalak egyike nem hímsteril, az önmegtermékenyítés megakadályozása költséges kézi munkát igényelhet. Továbbá, a hibrid vetőmagból kelt növények utódaiban a tulajdonságok hasadnak és az ún. hibridvigor elveszik. Ez azt jelenti, hogy a hibrid

vetőmagot minden évben újra elő kell állítani, azaz a hibrid növények magjai nem visszavethetőek. Amennyiben a hibrid genotípust az utódokban sikerülne fixálni, akkor nagyságrendekkel csökkenteni lehet a vetőmag előállítás költségeit és a termesztőnek, sem kellene minden évben új vetőmagot vásárolnia, ami főleg a fejlődő országokban könnyítené meg az emberek életét. Ez viszont nem érdeke a hibrid vetőmagot előállító és forgalmazó vállalatoknak.

A búza petesejt is képes megtermékenyítés nélküli (parthenogenetikus) fejlődésre, de az esetek nagyon kis számában indul el ez az egyedfejlődési út, ezért gazdasági jelentősége jelen pillanatban még nincs. Az ún. „Salmon” búza, amelyet 3 független (egy szexuális és két fertilizáció független embriófejlődésű) búza fajtából hoztak létre keresztezésekkel, viszont 90%-osnál nagyobb valószínűséggel választja a parthenogenetikus fejlődési utat. Sajnálatos azonban, hogy a termelékenységi mutatói nagyon rosszak ennek a fajtának. Az izolált „Salmon” típusú petesejtek *in vitro*, az eredeti szöveti környezetüktől elválasztva is, képesek megtermékenyítés hiányában a korai embriogenezis lépéseinek végigkövetésére. Ebből arra lehet következtetni, hogy az embriogenezis beindításához nem minden esetben szükségesek a petesejttől független külső jelek, hanem magában a petesejtben, illetve a zigótában kell keresnünk azokat a géneket, fehérje termékeket, amelyek felelősek a parthenogenezis/ apomixis folyamatának a beindításához (Naumova és Matzk, 1998).

2. 1. 3. A növényi megtermékenyítés molekuláris biológiája

Mivel növényekben a megtermékenyítés az anyai szöveteken belül történik, ezen folyamatok vizsgálatához elengedhetetlen a mikromanipulációs technikák magas szintű használata (Barnabás és mtsai., 2001). A molekuláris technológiák fejlődésével lehetővé vált, hogy akár egyetlen sejtből kiindulva is képesek vagyunk expressziós vizsgálatok elvégzésére (Brandt és mtsai., 2002; Okamoto és mtsai., 2005).

A megtermékenyítés az addig nyugvó petesejtet nyilvánvalóan aktiválja (anyagcsere, sejtfallszintézis, osztódás), kérdés azonban, hogy ez a kezdeti aktiváció együtt jár-e a petesejt saját génjeinek aktiválódásával. Az állatok esetében az ún. anyai/zigóta átmenet, vagyis a zigóta saját génjeinek a kifejeződése, fajtól függően késleltetve, csak a

zigóta 2-12 sejtes állapotában kezdődik meg (lásd Grimanelli és mtsai. 2005, és hivatkozásai). Ezt megelőzően az anya által a petében elhelyezett fehérjék és RNS-ek biztosítják az állati embriók fejlődését („anyai hatás”). Hogy hasonló mechanizmusok léteznek-e a növények esetében, sokáig nem volt egyértelműen megválaszolható, éppen technikai nehézségek következtében. Ugyanakkor az a tény, hogy a növények esetében megtermékenyítés nélkül és szomatikus sejtekből is fejlődhetnek embriók (lásd alább), jelzi, hogy az embriogenezis függetleníthető az anyai hatástól, legalábbis bizonyos esetekben.

Az utóbbi években elsősorban a lúdfű mutánsok segítségével végzett genetikai vizsgálatok igazolták az anyától örökölt allélek preferenciális kifejeződését a korai magfejlődés alatt (Grossniklaus és mtsai. 1998, Vielle-Calzada és mtsai. 2000) Ezt megerősítették kukoricával végzett kísérletek is (Grimanelli és mtsai. 2005). Ezek az eredmények azonban nem adnak tájékoztatást arról, hogy ezeknek a géneknek a transzkriptumait még az anyai szervezet helyezte el a petesejtben, vagy már a zigóta génműködését jelzik. A megtermékenyítéstől függő embriogenezis során a génexpresszió tanulmányozását nagymértékben megnehezíti a növényekben a megtermékenyítés kettőssége, azaz a petesejt és a centrális sejt egyidejű megtermékenyülése és az embrió valamint az endospermium fejlődésének párhuzamossága. Ennek megkerülésére két megközelítést alkalmaztak. Az egyik esetben a kukorica és az apomiktikus *Tripsacum* hibridjének embriófejlődését vizsgálták. Ez utóbbi esetben az embriófejlődés a megtermékenyítést megelőzően, az endospermium fejlődését több nappal megelőzve, következik be (Grimanelli és mtsai. 2005). cDNS microarray segítségével megállapították, hogy az embrió jelenléte nem változtatja meg szignifikánsan a génexpressziós mintázatot, az embriót nem tartalmazó embriózsákhoz képest. Ebből arra következtettek, hogy a zigóta genom aktiválódása nem szükséges az első embrionális osztódásokat megelőzően. A megközelítésnek azonban van egy nagy hátránya: az embriózsák sejtjeiből származó RNS mintákban az embrióra jellemző transzkriptumok alulreprezentáltak a szomatikus sejtek transzkriptumaihoz képest, így azok detektálása meghaladhatja a használt vizsgálati módszer érzékenységét. Ezt a hátrányt küszöböli ki a petesejtek és zigóták izolálása és sejtspecifikus cDNS könyvtárak készítése, amit a transzkript populáció random

szekvenálása és a szekvenciák analízise követ („EST” vagy „expressed sequence tag” analízis).

Ezt a stratégiát követték Sprunck és munkatársai (Sprunck és mtsai. 2005), akik összesen 1197 expresszáladó szekvenciát (EST) határoztak meg izolált búza petesejtekből és kétsejtes zigótákból. Nyilvánvalóvá vált, hogy a kétsejtes proembrió transzkriptum mintázata jelentősen eltér a petesejtétől, azaz ezek a kutatások a zigóta genomjának viszonylag korai (első sejtosztódás előtt vagy közvetlenül utána) aktiválódására utalnak.

Okamoto és munkatársai (2005) kétsejtes kukorica embrió apikális illetve bazális sejtjének génexpressziós mintázatát hasonlították össze, és jelentős különbségeket találtak. Azt is megfigyelték, hogy a csak a bazális vagy csak az apikális sejtre jellemző abundáns transzkriptumok nagy része már korábban, az egysejtes zigótában is, magas szinten jelen volt. Ez arra utalhat, hogy ezek a transzkriptumok a zigótában való aszimmetrikus elhelyezkedésük következtében kerülhettek át preferenciálisan az egyik utód sejtbe, de nem lehet kizárni az osztódást követő preferenciális mRNS degradációt sem. A gének többségének a funkcióját nem sikerült azonosítani, így szerepüket a sejtek további fejlődésének meghatározásában nem ismerjük. A sejtek sorsát meghatározó szerepe lehet azonban a lúdfű kétsejtes embriójának apikális illetve bazális sejtjeiben differenciáltan kifejeződő *wuschel homeobox2 (wox2)* illetve *wox8* géneknek (Haecker és mtsai. 2004).

2. 1. 4. A szemfeltöltődés folyamata

A megtermékenyítés utáni 10 napot szemnövekedési, az azt követő 20 napot szemfeltöltődési, majd a mag érésének utolsó 10 napját kiszáradási fázisnak hívjuk. A megtermékenyítést követő 4.-ik napra a mag mérete kb. 3-szorosára, növekszik, az embriózsákot körbeölelő anyai szövetek erősen megnyúlnak. Az embriózsákon belül kialakul a nagy központi vakolum, ami majd tápanyagraktárként funkcionál. Az embriózsák szélén elhelyezkedő citoplazmában gyors sejtmegosztódás történik, amit nem követ sejtfolképződés. A folyamathoz szükséges energiát az embriózsák tetején elhelyezkedő ún. antipodális sejtek biztosítják. A most megtörtént sejtmegosztódások száma erősen korrelál a későbbi endospermium sejtek számával és az érett mag tömegével. A zigóta ez idő alatt mindössze egyet, vagy kettőt osztódik és az embriózsák

alján helyezkedik el. A megtermékenyítést követő 4.-ik és 10.-ik nap között az endospermium sejtek tovább folytatják az osztódásukat, de ebben a fázisban már kezd kialakulni a sejtfaluk is. A 7.-ik napra kialakul az első összefüggő, gyűrű alakú, sejtfallal rendelkező endospermium réteg, az ún. szinciciális endospermium. Ebből a sejtrétegből alakul ki az aleuron réteg. A beljebb fekvő endospermium sejtekből, az ún. „keményítő” endospermium lesz, raktározási funkcióval. Az antipodális sejtek eltűnnek. Új struktúra alakul ki a nucelláris nyúlvány, amely a tápanyag továbbításban fontos szállítószöveteket tartalmaz. Az embrió fejlődése során kialakul a gyököcske és a hajtáscsúcs. Az embriót eddig tápláló, sűrű citoplazmatikus endospermium (celluláris endospermium) elkezd degradálódni. A szemfeltöltődés első 5 napján a szem még mindig nagyon lágy és az anyai szövetekben levő kloroplasztok miatt még zöldes színű. A merisztematikus endospermium sejtek egészen a 16.-ik napig osztódnak. A 16.-ik nap végére megjelennek az első „A” típusú nagyméretű keményítő, valamint fehérje és lipid szemcsék. A ventrális árokhoz közel levő aleuron sejtek speciális szállító sejtekké alakulnak, de a dorzális oldaliak továbbra is osztódnak, lehetővé téve a mag további növekedését. A zigóta tovább növekszik, és hosszúkás alakot vesz fel, a 16.-ik napra teljesen elfogyasztja a celluláris endospermiumot. Kialakul a szkutellum és az embrió, innentől kezdve ezen keresztül éri el a raktározott keményítőt. A következő 5 napban a mag elkezd sárgulni, és eléri maximális tömegét, amelynek jelentős része még víz, de az endospermium sejtek belseje már félig szilárd. Megjelennek a „B” típusú kisebb méretű keményítő szemcsék. A zigóta eléri teljes méretének a felét. A következő tíz napban a tápanyagok raktározásának sebessége eléri maximumát. Jelentős mennyiségű fehérje és keményítő raktározódik el ebben a fázisban. Az endospermium sejtek kezdenek elhalni. A zigóta eléri teljes fejlettségét és jelentős mennyiségű saját tápanyag raktárakat alakít ki, ami főleg tri-acil-glicerolokat és a kiszáradástól védő ozmoprotektív fehérjéket tartalmaz. Az utolsó tíz napban a mag folyamatosan vizet veszti és kiszárad. Az embrió és az aleuron sejtek nyugalmi állapotba vonulnak. (<http://www.wheatbp.net/cgi-bin/grain2.pl>)

2. 2. A szárazság és/vagy hőstressz hatása a növényekre

2. 2. 1. A szárazság stressz

A növények az evolúció során kifejlesztették azokat a mechanizmusokat, stratégiákat, amelyek képesé teszik őket arra, hogy szuboptimális, vagy kedvezőtlen külső körülmények között is sikeresen szaporodjanak. Termesztett gabonanövényeink esetében számunkra a növények túlélése nem fontos, de a termés mennyiségének és minőségének csökkenése igen. A két leggyakoribb stressz, ami a termék mennyiséget/minőséget csökkenti az a magas hőmérséklet és a vízhiány. Sajnálatos módon ez a két stressz gyakran együtt fordul elő, ami kedvezőtlen hatásait felerősíti. A hosszú távú éghajlati előrejelzések azt mutatják, hogy hazánkban (a Kárpát medencében) egyre gyakoribbak lesznek az ilyen nyarak.

A növények 3 különböző stratégiát fejlesztettek ki a szárazság káros hatásainak az elkerülésére:

- 1- a menekülés, amikor a reprodukció még a kedvezőtlen körülmények előtt bekövetkezik. Ezt rövid életciklussal, és/vagy a tápanyagok hatékony tárolásával érik el.
- 2 - az elkerülés, amikor a növények magas ozmotikus potenciált tartanak fenn. Ez lehet a vízvesztés minimalizálása (a sztomák zárva tartása, csökkentett levélfelület), vagy a vízpótlás maximalizálása (megnőtt gyökérzet).
- 3 - a tolerancia, amikor a növény képes elviselni a nagymértékű ozmotikus változásokat, és később kedvezőbb körülmények között rehidratálódik.

Azt, hogy melyik stratégiát választják a növények, függ a fajtól, az adott fajta genotípusától, a szárazság hosszától, mértékétől, a fejlődési állapottól és a szerv, sejt típusától (Rampino és mtsai., 2006; Araus és mtsai., 2002; Bartels és Souer, 2004; Zhu és mtsai., 2005; Verdoy és mtsai., 2004; Cominelli és mtsai., 2005). A búza esetében a szárazság a gyökerek növekedését gyorsítja, a gyökérzet méretét növeli. A gyökér ABA hormont termel, ami a szárban csökkenti a vízpoteenciált, sztóma záródást okoz, és

csökkenti a fotoszintézis hatékonyságát, ezzel csökkenti a tárolt energia (szénhidrát) mennyiségét, ami végül is csökkent növekedéshez (sejtosztódás és sejtmeignyulás mértékének csökkenése) vezet. A kevés tartalékolt tápanyag miatt a virágzáshoz és magfejlődéshez elengedhetetlenül szükséges energia nem áll rendelkezésre, ami pollen sterilitáshoz, vagy a magkezdemények abortációjához vezet.

A védekezés első lépése a kedvezőtlen külső jelek felfogása és feldolgozása. Arabidopsis növényen kimutatták egy transzmembrán hisztidin kináz fehérje receptor szerepét, ami más fehérje molekulákkal együttműködve felelős lehet az ozmotikus potenciál csökkenésének érzékelésében. A felfogott szignál különböző jelátviteli utakat aktivál. Ezeknek a jellemzői, hogy kinázok és foszfatázok játszanak fő szerepet benne és a kezdeti jelet felsokszorozzák. További szerepük, hogy biztosítják a stressz válaszok koordinálását, mivel az eltérő stresszhatások által aktivált jelátviteli utak egymástól nem teljesen függetlenek. A szervezetnek pedig lehetősége nyílik kombinált válaszokat adni az adott helyzetre. A jelátviteli folyamatok integrációjában fontos szerepet játszik a sejten belüli Ca^{2+} szint növekedése. A szárazság hatására számos Ca^{2+} -függő kináz aktiválódik. A jelátviteli láncok végső lépése, hogy transzkripciós faktorokat foszforilálnak/defoszforilálnak (aktiválnak, deaktiválnak), amelyek számos gén transzkripciójának szabályozását végzik. Ezek a gének ozmoprotektív és detoxifikálásban részt vevő fehérjéket, ion, víz és cukor transzportereket kódolnak. Lúdfűn végzett genom méretű transzkriptoszóma vizsgálatok több száz transzkripciós faktort azonosítottak, amelyek expressziója külső stressz hatására jelentősen változott. Ezeknek két fajtáját különítjük el: az 1. osztályba tartozó gének abiotikus stressz, a 2. osztályba tartozók mind biotikus, mind abiotikus stressz alatt regulálódnak. Ezen transzkripciós faktorok expressziós mintázata nagyon bonyolult és az általuk szabályozott stresszválaszban szerepet játszó gének expressziójának szabályozását egy hihetetlenül összetett hálózat végzi (Seki és mtsai., 2002, Shinozaki és mtsai., 2003, Sreenivasulu és mtsai., 2007, Chen és mtsai., 2002, Chen and Zhu, 2004, Zhou és mtsai., 2007).

Az egyre növekvő számú irodalmi adat lehetővé tette, hogy a szárazsághoz-kötődő transzkripciós egységeket definiálják. Ez az egység, a „regulon”, magában foglalja az adott transzkripciós faktort és az általa szabályozott géneket is. A DREB2 és NAC transzkripciós regulonok ABA függetlenül szabályozódnak (ezt az ABA inszenzitív abi1-

1 mutáns lúdfű vonal felhasználásával bizonyították), míg az AREB/ABF, DREB1, MYC és MYB transzkripció regulonok szabályozása ABA függő (Nakshima és Yamaguchi-Shinozaki, 2005).

A szárazságra olyan gének aktiválódnak, amelyek egyrészt megakadályozzák a sejtek további vízvesztését, segítik megvédeni a sejt struktúráját, fehérjéinek natív szerkezetét, valamint az esetleges sérüléseket, károsodásokat javítják. Fontos szerepet játszanak ebben a folyamatban a különböző cukor-transzporterek, amelyek mind a sejtmembránon, mind a tonoplaszton keresztüli cukor szállításért felelősek. A cukrok itt nem energiatároló szerepet töltenek be, hanem ozmolitként funkcionálnak. Az aquaporinoknak szerepe nem teljesen tisztázott, de feltételezhetően funkciójuk van a szárazság toleranciában (Alexandersson és mtsai., 2005). Fontos megemlíteni az ún. LEA (late embryogenesis abundant) fehérjéket, amelyek hidrofíl, globuláris fehérjék. Szerepük a sejtalkotók védelme kiszáradás során. Ehhez hasonló szerkezetű fehérjék termelődnek sóstressz esetén is (Shinozaki és Yamaguchi-Shinozaki, 1999).

2. 2. 2. A magas hőmérséklet hatása a növényekre

A magas hőmérséklet csökkenti a fejlődési szakaszok hosszát, felgyorsítja az érési folyamatokat. A rövidebb életidő miatt csökken a felfogott és hasznosított fény mennyisége, a szén asszimilációval kapcsolatos folyamatok, mint a párologtatás, fotoszintézis, légzés, zavart szenvednek (Stone, 2001). 35 °C fölött a Rubisco enzim aktivitása és a fotoszintézis hatékonysága nagymértékben csökken. Az optimálisnál 5 °C-kal magasabb hőmérséklet növeli az ABA szintézist, beindítja az antioxidánsok termelését, megváltoztatja a sejtorganellumok és a citoszkeleton szerkezetét és a membrán funkciókat is befolyásolja. A károsodás fő oka a membránfluiditás megváltozása, mind a kloroplasztban, mind a mitokondriummban, ami a fotoszintézis és a légzés megbomlott egyensúlyához, sejtkárosító oxidatív molekulák felszaporodásához vezet. A magas hőmérséklet az újonnan keletkezett fehérjék helytelen csomagolódását eredményezi, míg a már meglévő fehérjék esetén bekövetkezik a denaturáció. Ezen folyamatok megakadályozása a feladata a meleg hatására kifejeződő hő-sokk fehérjéknek (HSP), chaperonoknak. Ezek a fehérjék, a baktériumoktól a növényeken keresztül az emlősökig,

hasznos funkciót látnak el és szerkezetük is nagyon konzervált. Számos HSP bizonyítottan szerepet játszik lúdfű esetén a termotolerancia kialakításában (Burke és mtsai., 2000; Hong és Vierling, 2000), de gabona félék esetében még nem sikerült bizonyítani hasonló szerepüket. A HSP101 (ami a fő komponens a termotolerancia kialakításának lúdfűben), búza és kukorica homológjának vizsgálata kimutatta, hogy kifejeződése szerv és fejlődési állapottól függ, a mRNS és a fehérje szint változása nem mindig korrelál, ami arra utal, hogy transzkripcionálisan és poszttranszkripciósan is szabályozódik (Young és mtsai 2001). A lúdfű transzkriptoszómájának elemzése során a vizsgált 24000 gén 11%-ának az expresszióját jelentősen befolyásolta a hőstressz. Ezen gének között szerepelnek transzkripciós faktorok (21 gén, 3 osztály, 14 csoport, heat stress transcription factors (HSF)), a fehérje szintézisben, a jelátviteli folyamatokban, a metabolizmusban és a transzportban szerepet játszó gének (Busch és mtsai., 2005).

2. 2. 3. A magas hőmérséklet és a szárazság kombinált hatása a növényekre

A kombinált stresszhatások, mint a növénytermesztésben leggyakrabban előforduló szárazság és hőstressz hatására a válaszreakció nem a két különálló stresszre adott válaszok egyszerű kombinációja, hanem egy egyedi molekuláris, metabolikus reakció. Számos kísérlet és termesztési tapasztalat bizonyítja, hogy a kombinált hatás jóval jelentősebb, mintha csak egyedi stresszekkel kellene megküzdeni az adott növénynek, legyen az búza, árpa avagy kukorica (Heyne és Brunson, 1940; Craufurd és Peacock, 1993; Savin és Nicolas, 1996).

Gyakori, hogy a kombinált stresszre adott válaszok antagonisztikusak. Erre példa, hogy hőstressz hatására a légcserenyílások kinyílnak és a párologtatás révén hűtik a leveleket. Vízhány esetén a légcserenyílások záródnak, megakadályozva a túlzott folyadék veszteséget. A növénynek meg kell találnia azt az állapotot, ahol a „kecske is jól lakik és a káposzta is megmarad”, azaz a levelek se melegednek túl, de súlyos vízhány se alakul ki. A már sokat emlegetett lúdfű transzkriptoszómájának elemzése kimutatta, hogy hő és szárazságstressz együttes hatására egyrészt a két stresszre adott válasz részleges kombinációja figyelhető meg, de emellett 454 olyan gén expressziója változott meg

specifikusan, amely egyetlen egyedi stresszhatásra sem reagált. A kombinált stressznél a fő ozmoprotector nem a prolin, mint a szárazság esetén, hanem a szukróz. Hasonló eredményekre jutottak dohány esetében is, tehát elképzelhető, hogy ez egy általános jelenség a növények esetében (Rizhsky és mtsai., 2004).

2. 3. A szárazság és/vagy hőstressz hatása a búza ivaros folyamataira és a szemfejlődésre

2. 3. 1. A pollen és az embriózsák kialakulása stressz körülmények között

A búza esetében a kalászosítás előtti szárazság stressz drasztikus, akár 40 – 70 %-os termés csökkenést okozhat, amelynek elsődleges oka, hogy steril, funkcióképtelen pollenszemek alakulnak ki (Saini és mtsai., 1984). A belső pollenfal, ami pektocellulózból épül fel, nem tud normálisan kifejlődni, nem rakódik le elegendő tartaléktápanyag (keményítő) a pollenszem vegetatív sejtjének citoplazmájában. Ezek a rendellenességek gátolják meg a pollenszemek megtapadását a bibe felületén és a normális pollentömlő növekedést. A portokokban csökken a vízzeloldható szénhidrátok mennyisége és változik a savas invertáz enzim expressziója (Saini, 1997). A savas invertázok meghatározó szerepét a hím meiosis idején történő vízmegvonás hatására kialakuló pollen sterilitásban (Koonjul és mtsai., 2005) igazolták.. Búzából három invertáz cDNS-t izoláltak, melyek közül kettő (Ivr1, Ivr3) a sejtfal, egy pedig a vakuoláris invertáz enzimet (Ivr5) kódolja. Ezek a gének preferenciálisan, de nem kizárólagosan a portokokban expresszázódtak. Míg a vízhiány a meiosis idején szelektíven gátolta az Ivr1 és az Ivr5 gén transzkripcióját, addig az Ivr3-at nem befolyásolta. A meiosis kezdetétől a pollen éréséig tartó magas hőmérséklet stressz ($> 30\text{ }^{\circ}\text{C}$) csökkenti a pollenszemek életképességét, és ez a magkötés csökkenéséhez vezet (Saini és Aspinall, 1982). A pollen nem képes az összes HPS (heat shock protein) szintézisére, hőérzékenysége ezzel is magyarázható (Mascarenhas és Crone, 1996).

A tapasztalatok azt mutatják, hogy a női gametofiton nagyobb stressztűrő képességgel rendelkezik, mint a hím (Saini és Lalonde, 1998). Kukoricán végzett átfogó vizsgálatok (Moss és Downey, 1971) szerint az embriózsák kifejlődése során előidézett vízhiány miatt, a szárazság stressz mértékének és időtartamának függvényében, az

ovulumok 15-45%-a abnormálisan fejlődött, míg az kontrol növényeknél ez a jelenség nem haladta meg a 2,5%-ot. Az ABA szerepe nem tisztázott a kukorica szárazság stressz hatására bekövetkező termésesökkenésében (Ash és mtsai 2001). Megállapították, hogy a magházak ABA koncentrációja megnövekszik, de ez a különbség a virágzás idejére megszűnik a kontrol csoporthoz képest. Az asszimilátumok transzport folyamata korlátozott (a savas invertázok szerepe), a szénhidrát anyagcseréért felelős gének expressziója csökken, az öregedésért felelős gének kifejeződése nő a fiatal magházakban (Andersen és mtsai 2002; McLaughlin és Boyer, 2004). A magas hőmérsékletnek a búza női ivaros nemzedékre gyakorolt hatásáról is hiányosak az ismereteink. Az ausztrál Gabo búzafajtán tanulmányozták a meiosis idején alkalmazott magas (30 C°) hőmérséklet hatását (Saini és mtsai., 1983). A kezelés hatására a magházak egyharmada rendellenesen fejlődött. Előfordultak magházak, melyekben nem differenciálódott embriózsák és az üreget vagy a külső, vagy a belső integumentum sejtjei töltötték ki. A stresszhatásnak kitett termők 7%-ában a pollentömlők növekedése nem volt megfelelő. A szemkötődés átlagosan 21 %-kal csökkent.

2. 3. 2. Szemfeltöltődés hő és szárazság stressz alatt

A szemfeltöltődés idejének hossza és intenzitása az, ami meghatározza a búzatermés mennyiségét. Ez függ a keletkezett endospermium sejtek számától és azok „töltöttségi” szintjétől. A szemfeltöltődés első ciklusában jelentkező stressz elsősorban az endospermium sejtek számát, míg a későbbi stressz az endospermium sejtek méretét csökkenti, ami utóbbinak az oka a feltöltődési idő csökkenése és/vagy a feltöltődés intenzitásának a csökkenése. Ennek eredményeként csökken a keményítő és a fehérje tartalom, romlik a termés minősége, mennyisége (Setter és Flannigan, 2001). Búzában 20 napos vízhiány az endospermium kialakulása alatt 30-40% endospermium sejt szám csökkenést okozott és a nagyméretű „A” típusú keményítő szemcsék száma is 45%-kal csökkent (Nicolas és mtsai., 1985). A magas hőmérséklet úgyszintén gátolja a szemfeltöltődést valamennyi gabonafélében (Engelen-Eigles és mtsai., 2001; Wallwork és mtsai., 1998).

A szemfeltöltődés szoros kapcsolatban van az öregedési folyamattal és a szár tartalék tápanyagainak a mobilizálásával. A szárazság hatására gyorsul az öregedés, csökken a fotoszintetikus aktivitás, csökken a szemfeltöltődés hossza, de a tartalék tápanyagok magba áramlása nem gyorsul (Gebbing és Schnyder, 1999; Plaut és mtsai., 2004; Yang és Zhang, 2006; Dupont és Altenbach, 2003). A magas hőmérsékletnek hasonló hatása van: az érés meggyorsul, de a feltöltődés intenzitása is nő, ellentétben a szárazság stresszel. Ennek oka a metabolikus enzimek aktivitásának a hőmérsékletfüggése. Ha extrém magas hőmérséklet stresszt (37/28 °C) alkalmazunk a búza életciklusa 21 nappal kevesebb, mint a kontrol csoporté. A nagyobb méretű „A” típusú keményítőszemcsék aránya nőtt, a kisebb „B” típusúak aránya csökkent.

2. 4. Modern genetikai módszerek a növénynemesítésben

2. 4. 1. Molekuláris markerek

A gazdaságilag fontos gabonafélék esetén jó felbontású genetikai térképek léteznek, amelyek molekuláris adatokon alapulnak. A számunkra fontos tulajdonságok, legyen az a stressz rezisztencia, vagy a termés mennyisége, minősége, kialakítását nem egyedi gének, hanem inkább számos gén közös együttműködése határozza meg. Általában ezek a szorosan együttműködő gének egymáshoz közel, egy locuszon helyezkednek el. Ezeket nevezzük „QTL”-nek (quantitative trait loci, mennyiségi jellegek lokusza). A „QTL” analízisek fontosak, és jól használhatók árpa, rizs és kukorica esetében, de a búza hexaploid volta miatt alkalmazásuk nehéz a búzanemesítésben. Segítséget az nyújt, hogy a gabonaféléken belül a genetika felépítés, a gének sorrendje nagyon hasonló, így lehetőség nyílik más gabonafélékből (rizs, árpa) származó adatok extrapolálására is. Míg az abiotikus stresszek esetében számos „QTL” szerepét tisztázták (Langridge és mtsai., 2006), addig a sok próbálkozás ellenére, a szem minőségét döntő mértékben befolyásoló „QTL”-ek azonosítása eddig nem sikerült, feltehetőleg a változatos termesztési körülmények miatt (Bruce és mtsai., 2002). A QTL technológia azt is lehetővé teszi, hogy a termesztett és a nem házasított, vad fajták között meglevő eltérő tulajdonságokat (pl.

biotikus, abiotikus stressz rezisztencia) adott locuszokhoz, génekhez kössük (Langridge és mtsai., 2006).

2. 4. 2. Funkcionális genomika

2. 4. 2. 1. A transzkripció profil meghatározására szolgáló technikák

Fontos kérdés, hogy a gének milyen szövetekben, szervekben, fejlődési stádiumban mennyire szólnak meg. Az is fontos, hogy külső, vagy belső ingerek hatására hogyan változik meg expressziós szintjük. Az elmúlt évtizedekben számos módszer került kidolgozásra, ami alapján ezt a kérdéskört vizsgálni lehetett.

Northern blot: Az izolált RNS-t denaturáló gélen megfuttatják, majd membránhoz kötik és a radioaktívan, vagy később digoxigenin-nel, megjelölt próbával meghibridizálják. A filmen megjelenő fekete csík intenzitása arányos a felvitt specifikus RNS mennyiségével. Kontrollként általában a rRNS-eket alkalmazzák. A módszer hátránya, hogy egyszerre csak egy, esetleg pár gén (ha méretük jelentősen eltér) expresszióját lehet vele vizsgálni.

Nuclease Protection Assay: Alapja az S1 nukleáz (vagy hozzá hasonló enzim), azon tulajdonsága, hogy a kettős szálú nukleotid bontó képessége jóval kisebb, mint az egyes szálúé. Az izolált mRNS populációhoz adják a jelölt próbá(ka)t. A nem hibridizáló DNS-t, RNS-t ledegradálják. Itt a hibridizáció nem egy nagyméretű membránon zajlik le, hanem kis térfogatú folyadékban, emiatt az érzékenység jóval magasabb, mint a Northern-blot-té. A mintát poliakrilamid gélen megfuttatják, majd autoradiográfiásan vizsgálják. További előnye, hogy ha különböző méretű próbákat terveznek, egy kísérleten belül számos gén expresszióját tudják vizsgálni.

Nuclear-Run-On Assay: Alapja az, hogy a sejtmagizolálás folyamán a transzkripció pillanatszerűen leáll, mert a transzkripcióhoz szükséges ribonukleotid szubsztrátok eltűnnek a rendszerből. Később ez a folyamat újraindítható külső, jelölt ribonukleotidok adásával és az átírás befejeződhet. Utána a jelölt mintát általában membránhoz kötött cDNS-hez hibridizálják. Az autóradiográfiás jelek arányosak a sejten belül pillanatnyilag szintetizálódó gének számával. A módszer hátránya, hogy a

sejtmagizolálás bonyolult és nem mindig kivitelezhető. Előnye, hogy nagyon pontos képet ad az adott pillanatban szintetizálódó génekről, ami fontos lehet gyors transzkripcionális változások monitorozására.

PCR (polimeráz láncreakció): Az érzékenységet jelentősen növelte, amikor a meglévő technikákat kombinálták a PCR reakcióval. E technika alkalmas egy specifikus DNS szakasz megsokszorozására. Két, a kívánt DNS-hez komplementer, egymással szemben álló oligonukleotid, DNS polimeráz, megfelelő puffer és szabad nukleotidok szükségesek a reakcióhoz. Az eljárás általában 20-30 ciklusból áll és mindegyik ciklus 3 lépést tartalmaz. A denaturáció során a kettős szálú DNS szálak szétválnak magas hőmérséklet hatására. A kapcsolódási lépés során a primerek a DNS mintához tapadnak, majd a lánc hosszabbítás során a primereket meghosszabbítja. A PCR terméket agaróz gélen elválasztják, és mennyiségüket meghatározzák. A meghatározás pontossága nagyban függ attól, hogy a PCR reakciót az exponenciális fázisában állítják-e le. Az érzékenységet tovább lehet növelni, ha a gélben található DNS-eket átblottolják membránra, majd ott jelölt próbával meghibridizálják.

Real-Time PCR: Valós idejű PCR alkalmazásakor folyamatosan, minden ciklusban figyelik a keletkezett termék mennyiségét. A legelterjedtebb és legolcsóbb, amikor a kettős szálú DNS-be interkalálódó fluoreszcens festéket (SYBR Green) tartalmaz a reakciónk. A módszer előnye, hogy speciális és drága jelölt oligonukleotid nem szükséges, de hátránya az, hogy a nem specifikus termékekbe is beépülhet a festék, ezért az eredmények megbízhatóságának növelése miatt széles körű ellenőrzések szükségesek. Más módszerek a „Förster Rezonancia Energia Transzfer”-en (FRET) alapulnak. Ennek a lényege az, hogyha egymáshoz közel helyezkedik el egymáshoz egy fluoreszcens, fénykibocsátó molekula és egy az általa kibocsátott fény hullámhosszán elnyelő molekula, akkor a fényelnyelő meggátolja, hogy távolabbra is eljusson az emittált fény, de ha egymástól távolabb helyezkednek el, akkor a fluoreszcencia mérhetővé válik. A „Taqman” próbák a polimerázok 5' nukleáz aktivitását használják ki. A próba (DNS szekvencia) egyik végén kibocsátó, a másikon pedig elnyelő molekula csoport van. Ha a próba a PCR termékhez hibridizál, a polimeráz nukleotidokra bontja a próbát. Az elnyelő és kibocsátó molekula(most már) tértávolságba kerül, így a fluoreszcencia mérhető. Egyszerre akár több gént is vizsgálhatunk, ha a próbáinkra egymástól eltérő spektrumú fluorofór molekulát

tervezünk. A módszer nem igényel hosszas optimalizációt, de elég drága. A molekuláris „beacon”-ök esetében, hasonlóan a „Taqman” próbákhoz, a próba egyik végén kibocsátó, másik végén elnyelő csoport van. A kettőt egy hajtú szerkezet köti össze. A hajtú hurok régióján belül vannak a szekvencia specifikus nukleotid részlegek. A próbák a reakció végéig intaktak maradnak. Szabad állapotban közel van egymáshoz a fény kibocsátó és elnyelő molekularészlet, és így nincsen jel, de ha hibridizálnak a PCR termékhez, akkor tártávolságra kerülnek és mérhető a fluoreszcencia.

Macro, ill. micr array: Képesek egyszerre nagyon sok gén expresszióját vizsgálni. A macro array-ek kevesebb próbát tartalmaznak, nagyobb a pontok mérete és ebből következően kisebb a pontok sűrűsége, mint a micro array-eknél (1mm, ill. 200 µm; 64 próba/cm², 10000 próba/cm²; 10 ng DNS, 250 pg DNS próba mennyiség). Macro array-nél a mintákat radioaktívan jelölik meg és autoradiográfiás módszerrel értékelik ki. Micro array-nél miközben az RNS mintából cDNS-t készítenek, a szintetizálódó DNS láncba fluoreszcensen jelölt nukleotidok, vagy speciálisan módosított nukleotidokat építenek be, melyekhez később kötnek fluoreszcens molekulát. Ezzel a mintával hibridizálják meg a szilárd hordozóra felvitt DNS-eket. A DNS-ek eredetük szerint lehetnek PCR termékek vagy szintetizált oligonukleotidok. Régebben elterjedtek voltak a cDNS chippek, de jelentőségük napjainkra csökkent. Ennek oka, hogy mivel a próba szekvenciák túl hosszúak voltak ahhoz, hogy a hibridizálási és mosási lépésekkel az egymáshoz hasonló gének között ne következzen be jelentős kereszt kikötődés, képtelenség az egymáshoz nagymértékben hasonló gének egyedi expressziós mintázatát vizsgálni. A szintetizált oligonukleotidokkal ennél jóval nagyobb pontosságot tudnak elérni, napjainkban ezek az elterjedtek. Különböző cégek más és más stratégiát alkalmaznak. Az egyik megközelítés az, hogy az előre szintetizált oligonukleotidot (50-70 bp) egy speciális inkjet technológiájú nyomtatóval az üveg alapú hordozóra nyomtatják, vagy piezoelektro motoros technológiával akrilamid mátrixra viszik fel. Ebben az esetben a chip árát az oligonukleotid szintézis költsége határozza meg. Az „Affymetrix” cég fotolitográfiás eljárással 25 bp hosszúságú próbákat szintetizál közvetlenül a hordozó felszínén. Minden próbának van egy tőle egy nukleotiddal eltérő variánsa, a negatív kontrol. Általában 11 ilyen pár alkot egy próba csoportot, amelyek ugyanarra a génre vannak tervezve és kb. 50-60000 próba csoport van egy chipen. Itt a fő költség a litografikus lemezek előállítás, ami

irányítja, hogy melyik szintézislépésben, melyik nukleotid épül be a próbába. Ez a technológia optimális gyakran használt, nagy sűrűségű chipek előállításához, de nagyon drágává teszi a kis példányszámban értékesíthető chipeket. Az „Agilent” cég kombinálja a két előbbi megközelítést és in situ, a hordozón, szintetizálja a hosszabb (max. 60 bp hosszúságú) próba oligonukleotidokat. Itt minden egyes lemezre más és más próbákat lehet tervezni, hatalmas flexibilitást nyújtva. A ritkán vizsgált, vagy még nem teljesen ismert szekvenciájú fajoknál (pl. búza) ez optimális megoldás, mert egyrészt kis példányszámban előállítva is rentábilis, másrészt a nem tökéletesen működő oligonukleotidokat a következő kísérletbe könnyedén le lehet cserélni. Búzára jelenleg csak az Affymetrix cég által készített oligonukleotid microarray érhető el. Ez a chip 61127 próba csoportot tartalmaz, egyenként 11 próba párral. Mostanáig két publikáció jelent meg a felhasználásáról. Egyik esetben (Able és mtsai, 2006) a meiózist vizsgálták, míg a másik esetben gének lokalizációjának meghatározását végezték (Close és mtsai, 2007).

EST szekvenálás: A transzkripció vizsgálatok egyik speciális formája. Ilyenkor egy adott sejtpopulációból, szövetből, szervből vagy akár teljes növényből, élőlényből mRNS-et tisztítunk. mRNS-ből cDNS könyvtárat készítünk és abból véletlenszerűen kiválasztva szekvenálunk meg nagyszámú cDNS klónt. Az előforduló szekvenciák egymáshoz viszonyított arányaiból lehet következtetni a gének expressziójának mértékére. A másik előnye, hogy lehetővé teszi új gének azonosítását. Ez különösen fontos olyan esetekben, ahol a faj genomjának még nem ismert a szekvenciája. A legtöbb klónkönyvtár stresszelt mintákból készült, de gyakoriak a különböző fejlődési stádiumú alapanyagból készített könyvtárak is. Napjainkban millió fölött van az EST-k száma búzában, rizsben és kukoricában is. Mivel a búza genomja hexaploid és mérete is jelentős (a rizsnél 30-szor, a kukoricánál 8-szor nagyobb) ezért itt különlegesen fontos ez a megközelítés, mert, gazdasági okok miatt nem várható, hogy a közeljövőben ismert legyen a búza teljes genom szekvenciája. Napjainkra a DNS szekvenálási technológia fejlődése (Massively parallel sequencing, Rogers and Venter 2005), lehetővé teszi akár transzkriptómok gyors megszekvenálását is (pl. Weber et al. 2007).

2. 4. 2. 2. Genomikai megközelítések a gabonafélék stressztűrésének vizsgálatában

Az átfogó génexpressziós vizsgálatokra alkalmas elérhető macro- és microarray-ek száma folyamatosan nőtt az elmúlt években. Zhu és mtsai. (2003) 21000 próbát (a rizs genom kb. fele) tartalmazó rizs oligochipen vizsgálta a szemfeltöltődést. Előzetesen 491 olyan gént választottak ki, amelyeknek feltehetően köze van a szemfeltöltődéshez (pl. zsírsav illetve keményítő szintézisben részt vevő gének). Ezek közül 98 gén expressziója korrelált a szemfeltöltődéshez. Az expressziós profil hasonlósága alapján további 171 gént azonosítottak. Mivel a rizs teljes genom szekvenciája elérhető, lehetővé vált az *in silico* promóter analízis. Ennek során 9 olyan transzkripció faktorot találtak, amely ezeknek a géneknek a regulációjáért felelős. A magas hőmérséklet hatását a keményítő szintézisben részt vevő gének expressziójára (Yamakawa és mtsai., 2007) vizsgálták rizsen. Megállapításaik összecsengtek a már vizsgált biokémiai változásokkal (csökkent szem tömeg, kisebb amiláz mennyiség és az amilopektin láncok megnagyobbodása). A GBBS, SBE és az AGP-ázok kis alegységének expressziója csökkent, az α -amilázé pedig nőtt. Érdekes kísérlet volt, amikor két búzafajtát kereszteztek és az utódok közül különböző (magas illetve alacsony) „párologtatási hatékonyságú” vonalakat válogattak ki. Vízhányos körülmények között a párologtatási hatékonyság és a vízfelhasználási képesség összefügg a termésmennyiséggel. A 16000 vizsgált gén közül 93 mutatott eltérő expressziót a különböző vonalakban. Számos lehetséges regulátor gént sikerült azonosítani, ami lehetővé teszi, hogy a későbbi nemesítés során ezeket az adatokat felhasználva, nagyobb hozamú fajtákat lehessen előállítani.

Fehérje szinten végzett expressziós (proteomikai) vizsgálatok folynak minden jelentős gabonafélén. Hőstressz hatására búzában a keményítő szintézisben szerepet játszó fehérjék mennyisége csökkent, míg nem túl meglepő módon, a hősokk fehérjék mennyisége nőtt (Majoul és mtsai., 2003, 2004). 121 fehérjét azonosítottak búzában, melyeknek hőstressz hatására jelentősen megváltozott a mennyiségük (Hajheidari és mtsai., 2007). Ezeknek a két-harmada a thioredoxin célpontja. Ezen vizsgálatok lehetővé teszik, hogy meghatározzuk azokat a fehérjéket, melyek felelősek a különböző genotípusú fajták eltérő tulajdonságaiért.

2. 4. 3. Modern génebézészeti módszerek

A modern genetikai módszerek lehetővé teszik, hogy egyedi géneket juttassunk be, vagy üssünk ki a célorganizmus genomjában. Ez nagy lehetőséget nyújt arra, hogy egyedi, új tulajdonságokkal vértesszük fel a növényeket. A fő gabonafajták génszabványi úton történő módosítása technikailag ma már lehetséges (Shrawat és Lörz, 2006; Vasil 2007). Habár az Európai Unió belül a genetikai módosításokkal szemben jelentős politikai és társadalmi ellenérzés van, számuk és gazdasági jelentőségük világszerte folyamatosan nő. Mind lúdfűben, mind más növényi modell fajokban egyre több olyan gént azonosítanak, amelynek szerepe van a stressz tolerancia kialakításában. Ezek mind-mind lehetőségek arra, hogy javítsunk a növények termeszthetőségén (Zhang és mtsai., 2004; Nakashima és Yamaguchi-Shinozaki, 2005). Gabonafélék esetében is számos példa van már ennek a lehetőségnek az igazolására. A kukorica ASR1 gént (ABA és szárazság stresszelt növényekből azonosították mind mRNS, mind fehérje szinten) túlermelő vonalak fokozott öregedéssel válaszolnak a szárazságra. Oka, hogy ez a feltételezett transzkripciós regulátor megváltoztatja az anyagcsere rátáját a felhasználó (sink) és termelő (source) szervek között és szánhidrát újraelosztást indukál. A foszfo-enol-piruvát-karboxiláz túlermelése javította a vízfelhasználás hatékonyságát és mérsékelt szárazság hatására a kontroll csoporthoz képest magasabb szárazanyagtartalmat produkált (Jeanneau és mtsai., 2002). Az árpa HVA1 gént túlermeltetve búzában a genetikailag módosított vonalaknak megnőtt a stressz toleranciájuk, a levelek víztartalma és a termés mennyisége magasabb volt, mint a kontroll csoporté (Sivamani és mtsai., 2000, Bahieldin és mtsai., 2005). Hasonló eredményeket kaptak rizs esetén is (Xiao és mtsai., 2007).

Az elkövetkező években várhatóan számos eddig még nem vizsgált gén hatását fogják megvizsgálni különböző kutatócsoportok. Várhatóan a jelátviteli folyamatokban résztvevő kinázok/foszfátázok, transzkripciós faktorok, ozmoprotektáns és más stresszhez köthető gének szerepét fogják elsősorban górcső alá venni.

3. Célkitűzések

Munkám célja a búza transzkriptószómájának bioinformatikai és molekuláris biológiai vizsgálata, és azon gének kiválasztása, amelyek fontos szerepet játszanak a megtermékenyítés és a szemfeltöltődés folyamatában. Ennek a célnak az érdekében az alábbi feladatokat tűztem ki magam elé:

- egy széles körben alkalmazható, olcsó, búzaszem specifikus géneket reprezentáló oligonukleotid chip előállítása, melyet a szelekciós folyamatok során a nemesítők is használhatnak.
- egy , a búzaszemben expresszáldó gének nagyrésztét reprezentáló, elsősorban alaputatásban használható oligonukleotid chip a szemfeltöltődés és a szemfejlődés folyamatainak a tanulmányozásához.
- EST könyvtárak létrehozása pete ill., fiatal, még nem osztódott zigótából és cDNS-ek random szekvenálása során pete ill., zigóta specifikus gének meghatározása. Ezzel lehetővé válhat a gének promóterjeit izolálva csak petében és fiatal zigótában expresszáldó génkonstrukciók elkészítése (apomixis/parthenogenezis kiváltása).
- A nagymennyiségű szekvencia adat elemzését megkönnyítő adatbázis és szoftver elkészítése.

4. Anyagok és módszerek

4. 1. Programozás és bioinformatika

4. 1. 1. A programok elkészítéséhez használt fejlesztői környezet

A munkám során elkészített szoftvereket Microsoft Visual Basic.NET 2005-ös és 2008-as kiadásának „Express Edition” verziójú keretrendszerével végeztem. Az adatbázis kezeléséhez a Microsoft Sql Server 2005-ös „Express Edition” verziójú szoftverét használtam. Mindkét szoftver ingyenes, feltéve, ha nem kereskedelmi célokra használják. Az ingyenes verziók esetén nem elérhető minden szolgáltatás. Például a maximálisan kezelhető adatbázis mérete 4 Gb, amit jelen esetben csak az adatok tömörítésével, vagy az adatok több adatbázisba való szétosztásával tudtam elérni. Mivel munkám során törekedtem, hogy a megírt program modulok később egy esetlegesen piacon is elérhető termék részei lehessenek, nagymértékben optimalizáltam a működésüket. Az alábbi névtereket, osztálykönyvtárakat és a benne foglalt függvényeket, metódusokat használtam.

System.IO Fájlműveletek. Adatok beolvasása és elmentése. Használat során arra törekedtem, hogy az írás, olvasás során ne egész fájlokkal, hanem csak egy részükkel dolgozzam (stream, folyam technológia), hogy a nagy méretű fájlokkal végzett műveletek ne terheljék le túlságosan a háttértárolókat és a memóriát.

System.Windows.Forms Ez az névtér tartalmazza a legtöbb felhasználó számára ismerős vezérlőt (szöveg beviteli mező, lenyúló lista, parancs gomb, menü, táblázat megjelenítő, falista)

Microsoft.SqlServer.Management.Smo A névtéren belüli osztályok és azok függvényei és metódusai használatával az SQL server 2005-en keresztül lehetséges a gépen található adatbázisok tulajdonságainak lekérdezése, módosítása és létrehozása. Ezen keresztül valósítható meg az adatbázis fájlokról biztonsági mentés készítése, vagy ezekből való visszaállítása.

System.Data.SqlClient Feladata az adatbázisokból adatok lekérdezése és feltöltése.

System.Net Az internetes kommunikáció ezen keresztül valósul meg.

System.Text A szöveg típusú változók kezelésére való.

StringBuilder osztály: Használatával lehetséges előre meghatározni a szöveg típusú változó méretét és a memoriába előre lefoglalni a megfelelő mennyiségű területet. Fájlok ki és beolvasása közben nagyon hasznos.

RegularExpressions: Minták felismerésére képes. Főleg a html fájlok feldolgozása során használtam.

Ezen kívül a hierarchikus adatok megjelenítésére a Developer Express cég XtraTreeList vezérlőjét használtam. A vezérlő próba verziója ingyenes, csak felbukkanó ablakban figyelmeztet erre, valamint kereskedelmi forgalomba nem hozható.

4. 1. 2. Az adatok beolvasása

Minden beolvasáskor lehetőség van egy vagy több, vagy pedig egy mappában található összes fájl beolvasására. A szoftver a következő fájltypusokat képes értelmezni:

FASTA: szekvencia leíró fájl. Az egységek első betűje mindig ">". Utána következik a szekvencia azonosítója majd, az első szóköz után az annotáció. A következő sortól kezdve a szekvencia kezdődik és tart a fájl végéig, vagy a következő „>” jelig.

GenBank: A legelterjedtebb szekvencia leíró fájltypus. Rengeteg lehetséges annotációval. Jelen pillanatban a szoftver csak az általunk fontos annotációkat keresi és dolgozza fel, de későbbi fejlesztés során majd az összes annotációra elérhető lesz a szűrés

és feldolgozás. A kezdő sor mindig „LOCUS” szóval kezdődik és mindig egy „chr(10) chr(13) //” végződik.

BLAST: A BLAST program kétfajta eredményét képes beolvasni. Az m9 (tabulátorral tagolt) forma egy tömör, kis méretű és csak a legfontosabb adatokat tartalmazza. Az m7 (XML) forma a homlóg gén annotációját és a szekvencia összehasonlítást (alignmentet) is tartalmazza. Egy példán keresztül: a petéből származó 243 gén elemzése során az m9-es fájl 1,5 Mb, az m7-es fájl 38 Mbyte volt.

Tabulátorral tagolt fájlok. A különböző szoftverek között ez az egyik leggyakoribb adatcserélő forma, bár mostanába kezdi leváltani a XML fájltypus. Munkám során számos ilyen fájlt kellett adatbázisba importálnom, de mindegyik egy kicsit másmilyen volt.

a,	assembly_accession	genbank_accession
	TA49712_4565	DY543273
	TA49712_4565	BJ218830
	TA49712_4565	BJ272153
	TA49712_4565	CD932062

Ez a legegyszerűbb elrendezés, semmit se kell módosítani rajta. Ez „egy a sokhoz” kapcsolatot ír le, ahol az első oszlop egyes elemeihez a második oszlop több eleme is tartozhat.

b,	assembly_accession	protein_accession
	TA49712_4565	UniRef100_Q84KI0
	TA49713_4565	UniRef100_Q8W430
	TA49714_4565	UniRef100_Q4AEI9
	TA49715_4565	UniRef100_Q4AEI8

Ebben az esetben a második oszlopban az „UniRef100_” szövegrészlet nem hordoz információt, felesleges. Az adattáblák közötti kapcsolat kialakítását is lehetetlenné teszi, ha az azonosítók nem „egy az egyben” feleltethetőek meg egymásnak, ezért a második sorból ezt a részletet el kellett távolítani.

c,	#KO	GO
	K00001	[GO:0004022 0004023 0004024 0004025]
	K00002	[GO:0008106]
	K00003	[GO:0004412]

K00005 [GO:0008888]

Itt a második adatsorból el kell távolítani a „[“, “],”GO” karaktereket. A maradékot szünet karakterenként részekre bontani. Az így kialakuló „egy a sokhoz” típusú adattáblát vittem be az adatbázisba.

Mivel a fentiek számos kombinációja megfigyelhető ezért egy általános beolvasó vezérlőt készítettem (4. 1. ábra).

name	Remove	Subsplit	Indexing
#KO			Indexed
GO	[,].GO:		Unique
*			

#KO	GO
K00001	[GO:0004022 0004023 0004024 0004025]
K00002	[GO:0008106]
K00003	[GO:0004412]
K00005	[GO:0008888]
K00006	[GO:0004367]
K00007	[GO:0007813]
K00008	[GO:0003939]
K00009	[GO:0008926]
K00010	[GO:0050112]
*	

4. 1. ábra A tabuláris fájlok beolvasásáért felelős vezérlő

A tulajdonságai:

- ki lehet választani, hogy melyik adatbázisba kerüljön (TAGI_WHEAT_10)
- meg lehet adni az új adattábla nevét (KEGG_KO+GO)
- meg lehet adni, hogy az első sor leírás, vagy már adat, és be lehet állítani, hogy mik legyenek a mezők nevei az adattáblán belül.
- be lehet állítani, hogy ezek közül melyik mező egyedi, indexelt vagy egyszerre egyedi és indexelt.
- az eltávolítandó karaktereket vesszővel elválasztva kell megadni.
- bár a képen nem látszik, de egy szünetjel van begépelve a subsplit mezőbe.

A végeredmény az (4. 2. ábrán) látható.

	#KO	GO
▶	K00001	0004022
	K00001	0004023
	K00001	0004024
	K00001	0004025
	K00002	0008106
	K00003	0004412
	K00005	0008888
	K00006	0004367
	K00007	0047813
	K00008	0003939
	K00009	0008926
	K00010	0050112
	K00011	0004032
	K00012	0003979

4. 2. ábra A „KO” és „GO” azonosítók közötti kapcsolatot jelképező adattábla (részlet).

A program automatikusan megállapítja, hogy milyen adattípussal a leghatékonyabban kezelni az adatokat. A „#KO” mező esetén ez egy 6 jegyű szöveg, míg a „GO” mező esetén egy integer típusú szám. Ezeket a szoftver az első 10000 adatsor beolvasása alapján számolja ki. Ha a későbbi sorok ettől eltérő jellegűek, akkor automatikusan módosul az adattábla szerkezete.

Mivel előfordul, hogy nagy mennyiségű adatot kell egyszerre felvinni az adatbázisba, ezért az „Sql_BulkCopy” metódust használtam. Ez alkalmas nagy mennyiségű adat gyors bevitelére, de hátránya az, hogyha valahol hiba van (nem megfelelő adat típus, vagy nem egyedi érték, ahol azt beállítottuk), akkor az egész beolvasást törli, ami az esetek jelentős részében 100000 adatsor. Ezért a program mindig előzetesen leellenőrzi, hogy megfelelő-e a felviendő adat.

Uniref: Egy Uniref egység első sorának első két betűje mindig az „ID” és a záró sora mindig „/”. Minden sor első két betűje határozza meg az adattípusát. A teljesség igénye nélkül (részletesen: <http://www.pir.uniprot.org/start/faq.shtml>):

- a, ID egyedi azonosító
- b, DT a record felvitelének ideje
- c, DE leírás
- d, OS a faj azonosítója
- e, DR más adatbázisokhoz való kapcsolatot írja le (GO,KEGG,TIGRFAM stb.)

Mivel a teljes adatbázis mérete meghaladta a 10 Gbyte-ot, ezért csak azokat az recordokat vittem fel az adatbázisba, ami már szerepelt a TA-k annotációjánál.

Html fájlok: Itt minden egyes típusra specifikus belovásó modult kellett írni. Erre egy példa. A TIGR „Gene Index 10” búza szekciójának feldolgozása.

a, keresd meg az összes olyan szövegrészletet, ahol előfordul a „/cgi-bin/tgi/est_report.pl?gudb=wheat&EST=” karakterlánc.

b, keresd meg a következő „>” és a következő „>” karaktereket. A kettő között levő szöveg a könyvtár neve.

c, keresd meg a következő szöveget „<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=search&doptcmdl=genbank&db=nucleotide&term=>”. Ennek a vége és a következő aposztróf jel között van az EST azonosítója.

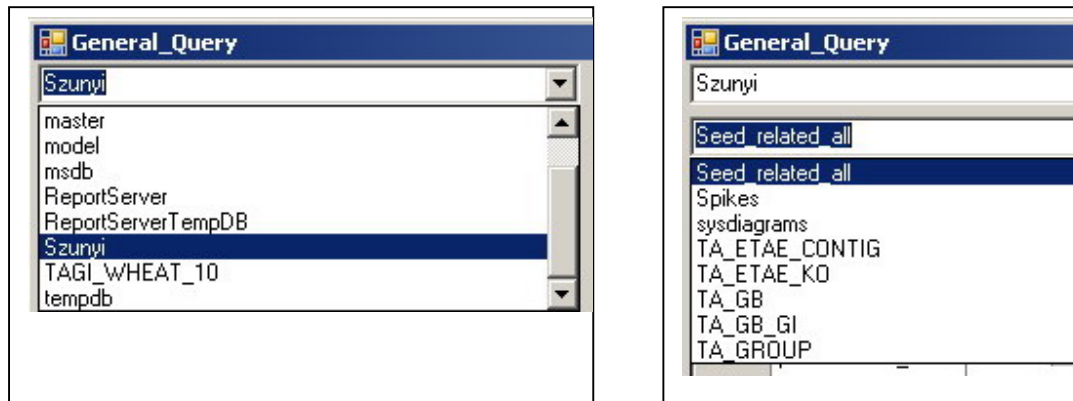
d, a „GO” azonosító megtalálásához keresd meg a „<http://www.godatabase.org/cgi-bin/amigo/go.cgi?query=>” szövegrészletet. A következő 10 karakter lesz a „GO” azonosító.

e, a génhez tartozó KEGG útvonal megtalálásához keresd meg ezt a szövegrészletet „map=map”. Ennek a vége és a következő aposztróf jel között van a pathway azonosítója.

4. 1. 3. Adatbázis műveletek

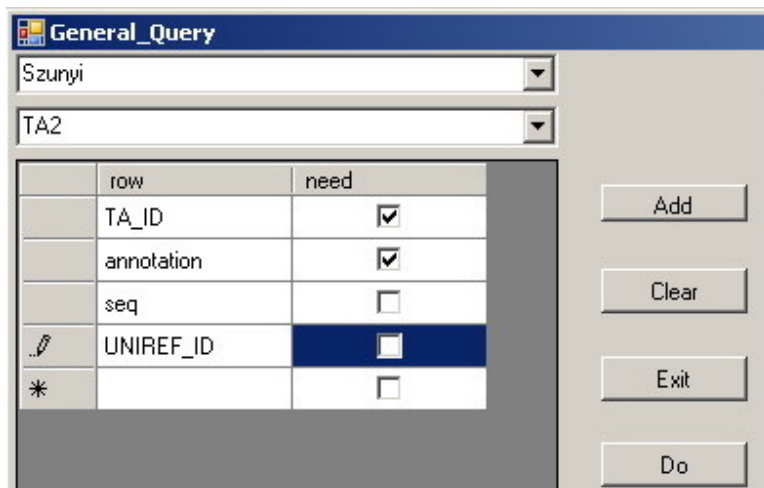
Számos gyakran használt lekérdezés „be van drótozva” a programba, de nem lehet előre látni, hogy milyen adatokkal, milyen struktúrában fogunk dolgozni később. Azon felhasználók számára, akik nem ismerik az SQL nyelv szabályait, írtam egy SQL tervező modult.

Meg kell adni, hogy melyik adatbázis, mely adattábláinak adataira vagyunk kíváncsiak (4. 3. ábra).



4. 3. ábra Az „Sql lekérdező tervező” modul részlete. Az adatbázis és adattábláinak a kiválasztása.

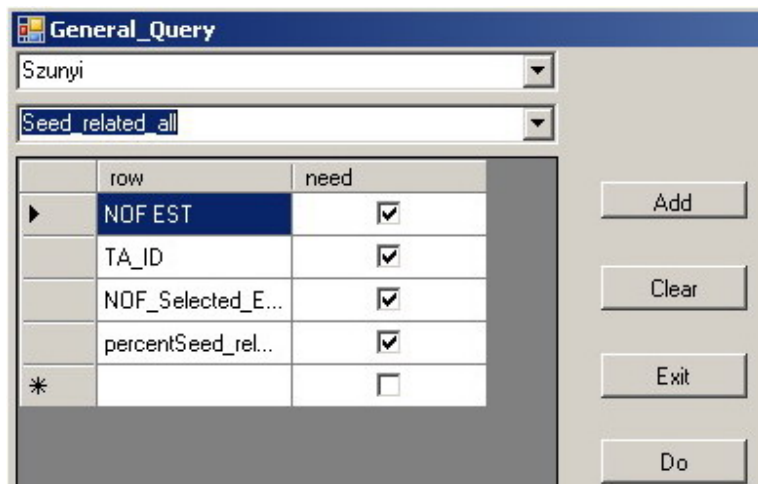
Annyi adattáblát adunk a kereséshez, amennyit akarunk. Minden adattáblánál bejelöljük azokat a mezőket, amikre kíváncsiak vagyunk. Jelen esetben a „Szunyi” adatbázis TA2 táblájából kell az azonosító és az annotáció. Az „Add” gomb lenyomásával hozzáadtuk a kereséshez (4. 4. ábra, 4. 5. ábra).



4. 4. ábra A lekérdezendő mezők kiválasztása a „TA2” adattáblából.

A „Seed_related_all” táblában tároltam, hogy melyik TA-t:
 - hány EST-ből szerelték össze,

- ezek között mennyi EST köthető a szemhez,
- ezeknek az EST-knek mekkora az aránya az összeshez képest.



4. 5. ábra A lekérdezendő mezők kiválasztása a „Seed_related_all” adattáblából

Az összes lekérdezendő mezőt tartalmazza a lenyúló lista, és mindegyik mezőre lehet szűrni (4. 6. ábra). Az elfogadott operátorok:

- > nagyobb
- >= nagyobb egyenlő
- < kisebb
- <= kisebb egyenlő
- = megegyező
- consist of tartalmazza
-

	Features	operator	Value
	[Seed_related_all].[NOF EST]	>	2
▶	[Seed_related_all].[percentSeed_related_al]	>	0.4
*	[TA2].[TA_ID] [TA2].[annotation] [Seed_related_all].[NOF EST] [Seed_related_all].[TA_ID] [Seed_related_all].[NOF_Selected EST] [Seed_related_all].[percentSeed_related_all]		

4. 6. ábra A szűrendő mezők és a küszöbértékek megadása

Az összetett kereséseknél meg kell adni, hogy az adattáblák melyik mezőnél fogva csatlakoznak egymáshoz.

	Features1	Features2
▶	[TA2].[TA_ID]	[Seed_related_all].[TA_ID]
*		

4. 7. ábra A közös azonosító megadása, ami összekapcsolja az adattáblákat

A software előállítja a lekérdezést, ami jelen esetben így néz ki.

```
Select [TA2].[TA_ID],[TA2].[annotation],[Seed_related_all].[NOF_EST],
[Seed_related_all].[TA_ID],[Seed_related_all].[NOF_Selected_EST],
[Seed_related_all].[percentSeed_related_all]
FROM [TA2] INNER JOIN [Seed_related_all] ON
[TA2].[TA_ID]=[Seed_related_all].[TA_ID]
Where ([Seed_related_all].[NOF_EST]>2) AND
([Seed_related_all].[percentSeed_related_all]>0.4)
```

Az eredmény egy adatábla, amit vagy az adatbázisba lehet felvinni, vagy pedig elmenti tabulárisan tagolt fájlba (4. 8. ábra). A software támogatja a többszörös csatolásokat is.

TA_ID	annotation	nof_EST	nof_Selected_EST	Percent_of_Selected_EST
TA49753_4565	Glycine rich protein-like [Solanum tuberosum (Potato)]	13	7	0.538462
TA49754_4565	Dbj BAA90612.1 [Arabidopsis thaliana (Mouse-ear cress)]	4	3	0.75
TA49751_4565		13	9	0.692308
TA49750_4565	Putative 24 kDa seed maturation protein [Oryza sativa (japonica cultivar-group)]	3	2	0.666667
TA49749_4565	OSJNBa0006B20.1 protein [Oryza sativa (japonica cultivar-group)]	43	34	0.790698
TA49747_4565		57	25	0.438596
TA49746_4565	Manganese superoxide dismutase [Triticum aestivum (Wheat)]	26	15	0.576923
TA49744_4565	putative auxin-induced protein [Oryza sativa (japonica cultivar-group)]	3	2	0.666667
TA49743_4565	Putative inorganic pyrophosphatase [Oryza sativa (japonica cultivar-group)]	7	4	0.571429

TA49742_4565	Cytochrome [Cichorium intybus (Chicory)]	66	36	0.545455
TA49741_4565	MADS-box transcription factor 13 [Oryza sativa (Rice)]	54	40	0.740741
TA49740_4565		5	5	1
TA49739_4565	Succinate dehydrogenase subunit 4 [Oryza sativa (Rice)]	8	4	0.5
TA49737_4565	Putative GTP-binding protein [Oryza sativa (japonica cultivar-group)]	22	16	0.727273
TA49736_4565		53	52	0.981132
TA49735_4565	Aspartic proteinase [Triticum aestivum (Wheat)]	58	32	0.551724
TA49734_4565	OSJNBa0019K04.9 protein [Oryza sativa (japonica cultivar-group)]	9	8	0.888889
TA49730_4565	Hypothetical protein [Arabidopsis thaliana (Mouse-ear cress)]	6	3	0.5

4. 8. ábra A keresés végeredménye (részlet).

Az eredménytábla 22287 sort tartalmaz. A lekérdezés futási ideje kevesebb, mint 5 másodperc.

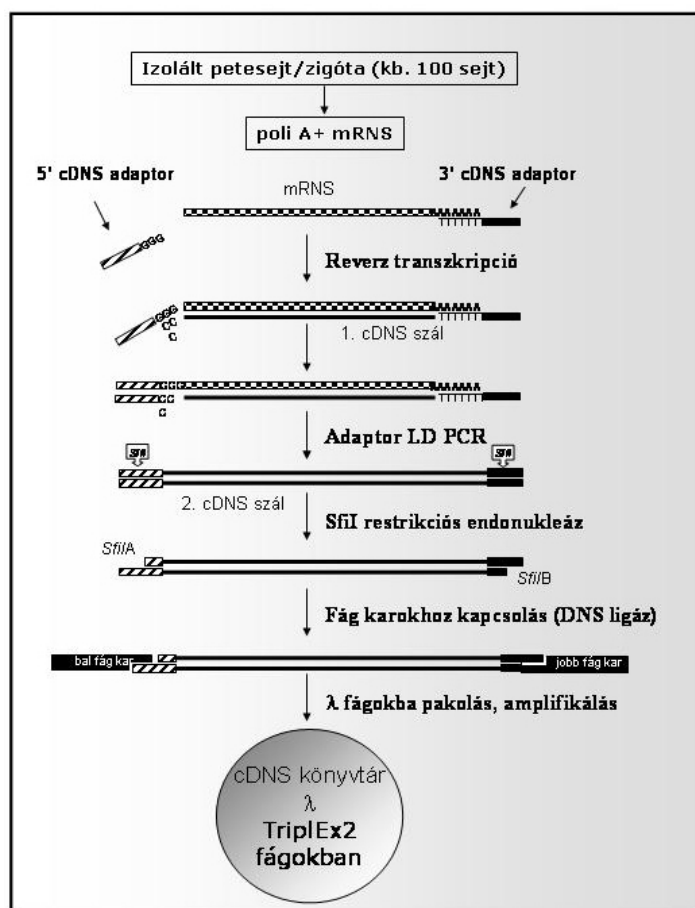
4. 2. A klónkönyvtárak előállítása

A sejteket a Dynal cég mRNA izoláló Kit-jének izoláló oldatában való inkubálással tártuk fel. A sejtmintákból az mRNA populációt közvetlenül izoláltuk mágnesezhető szemcsékhez kötött oligo-dT oligonukleotidokhoz történő molekuláris hibridizáción keresztül. Mágnes segítségével közvetlenül izoláltuk a szemcsékhez kötött tiszta mRNA populációt, lényeges veszteség nélkül, ami ilyen kis sejtminták esetében elsőrendű fontosságú. A tisztított mRNA-eket néhány (3-5) mikroliter RNáz mentes vízbe eluáltuk, így a teljes mennyiséget fel tudtuk használni a cDNS szintézishez. A kis mennyiség miatt a mRNA-eknek sem mérete, intaktsága (minősége) sem mennyisége nem volt meghatározható.

A cDNS szintézist és a cDNS könyvtár készítést a Clontech cég Kit-jei segítségével végeztük, amelyek az ún. SMART technológián alapulnak. Ez bizonyos (RNázH minusz pont mutáns) reverztranszkriptáz enzimek azon tulajdonságán alapul, hogy a mRNA molekula 5' végét elérve képesek láncot váltani, és így egy megfelelő oligonucleotidot a cDNS 5' végéhez kapcsolni. Ennek köszönhetően a transzkripciót elindító 3' cDNS szintézis primeren kívül az 5' végen is rendelkeznek a cDNS molekulák egy ismert szekvenciával, s így PCR segítségével felszaporíthatóak és irányítottan

klónozhatóak megfelelő vektor molekulába. A cDNS szintézis és könyvtár készítés főbb lépéseit a 4. 9. ábra szemlélteti.

A cDNS második szálát PCR segítségével szintetizáltuk a kit leírásait követve. Mivel azonban a leírtaknál sokkal kisebb mennyiségű mRNS-ből indultunk ki, a szükséges ciklus számot empirikusan kellett meghatároznunk. Eredményeink alapján kb. 50 sejtől származó mRNS-ből kiindulva mintegy 33-36 PCR ciklus szükségeltetik megfelelő mennyiségű kettősszalú cDNS előállításához. A kettős szálú cDNS az az első termék, amely a korábbi lépések sikerességére utal, hiszen ekkor rendelkezünk először vizsgálható mennyiségű anyaggal. Sajnos több esetben a sejtmintákból kettős szálú cDNS-t szintetizálni nem sikerült, ami utalhat mRNS degradációra illetve az első szál szintézisének sikertelenségére.



4. 9. ábra PCR alapú, néhány sejtől kiinduló, cDNS könyvtár készítés menete

Az előkísérletek során számos sejtmintát fel kellett használnunk a módszer paramétereinek optimalizálására. Két mintával jutottunk el a cDNS bank készítéséig. Érett petesejtekből és fiatal zigótából származó minta esetében sikerült megfelelő minőségű és mennyiségű cDNS-t előállítani. Ezeket SfiI enzimmal emésztettük, majd méret alapján frakcionáltuk. A megfelelő méretű frakciókat összekevertük és a cDNS populációt SfiI emésztett Lambda TriplEx2 fág karokhoz ligáltuk, három különböző (1:1, 1:2, 1:3) arányban. A ligátumokat fágokba pakoltuk és a fágokat megtitráltuk. Mindhárom fág kar:cDNS arány esetében igen jó hatékonysággal kaptunk plakkképző fágokat (pfu). Összességében milliliterenként mintegy 3×10^6 független fágot kaptunk, amely az mRNS populáció igen jó reprezentációjának tekinthető ($1-2 \times 10^6$ pfu/ml már annak számít), bár a PCR lépés miatt a redundancia foka nem becsülhető meg ebben a szakaszban.

A fág növesztést IPTG-t és X-Gal-t tartalmazó lemezekben is elvégeztük, amely lehetővé tette az α -komplementáció segítségével az üres illetve cDNS inszertet hordozó fágok arányának meghatározását. Ez a pete esetén kb. 20:80 %-nak adódott, amely jó arány, különösen a kiindulási anyag kis mennyiségét tekintve, fiatal zigótánál 40:60%. A cDNS könyvtárát pete esetében felamplifikáltuk, úgy, hogy 1.5×10^6 egyedi fágot szélesztettünk ki 26 db 20 cm-es átmérőjű lemezre (ennél több fizikailag nem kezelhető). A fágokat lemostuk, fág hígító pufferben felvettük és kloroformmal extraháltuk a bakteriális szennyeződés eltávolítására. A tiszta fág oldat (kb. 300 ml) képezi a tényleges fág cDNS könyvtárát. Ezt ismét megtitráltuk. Ezek alapján a cDNS könyvtárunk $> 3 \times 10^9$ pfu/ml fágot tartalmaz. A könyvtár nagy részét 7% DMSO-val kiegészítve lefagyasztottuk és -80°C -on tároljuk.

A cDNS könyvtár további jellemzésére, húsz fágból PCR segítségével felszaporítottuk a cDNS inszerteket és gélen ellenőriztük a méretüket. A cDNS-ek változatos hosszúságúak és van közöttük számos 1 kb feletti méretű is, ami a cDNS szintézis megfelelő minőségére utal.

A kiválasztott fágokat plazmiddá konvertáltunk. Ez a Lambda TriplEx2 fág esetében a megfelelő baktérium törzs megfertőzését és ezt követő szelekciót és plazmid DNS tisztítást jelent. A kapott plazmidokat a Macrogene céggel szekvenáltattuk meg.

A fiatal zigótából egyelőre nem tisztázott okból, nehezebben és kevesebb cDNS-t tudtunk előállítani, mint a petesejtekből. Ennek oka lehet pl. a sejtfal kialakulása és a feltáródás kisebb hatékonysága.

4. 3. RT-PCR

4. 3. 1. cDNS szintézis

A növényi szövetekből először RNS-t izoláltunk. A különböző búza szervekből vett szövetminták esetében a növényi anyagból először TRI-reagenssel össz-RNS-t tisztítottunk, és ezt követően ebből izoláltunk mRNS-t oligodT-vel fedett vasgyöngyök és mágnes segítségével (GenoPrep™ Direct mRNA Kit, Genovision). A többi kísérletnél a feltárt szövetekből közvetlenül mRNS-t tisztítottunk. A reakciókörülmények az alábbiak voltak: A RNS mintákhoz (1-3 µg) hozzámértük a CDS primert (oligo-dTimidin, 10 pmol) és a vizet, és az így kapott elegyet (11 µl) két percig 70 °C-on inkubáltuk, majd rövid időre visszahelyeztük jégre. A mintákhoz hozzámértük a pufferből (4 µl 5x RT-puffer), dNTP-ből (1 µl, 10 mM), RNáz inhibitoros vízből (1 µl, 100 egység) és RevertAid™ M-MuLV reverz transzkriptáz enzimből (1 µl, 200 egység, Fermentas) összeállított keveréket, majd egy órán keresztül 42 °C-on inkubáltuk, miközben az enzim megsztetizálta a kívánt cDNS-t. A reakciót felmelegítéssel (95 °C, 5 perc) állítottuk le.

4. 3. 2. Valós idejű kvantitatív reverz-transzkripció PCR (RT-QPCR)

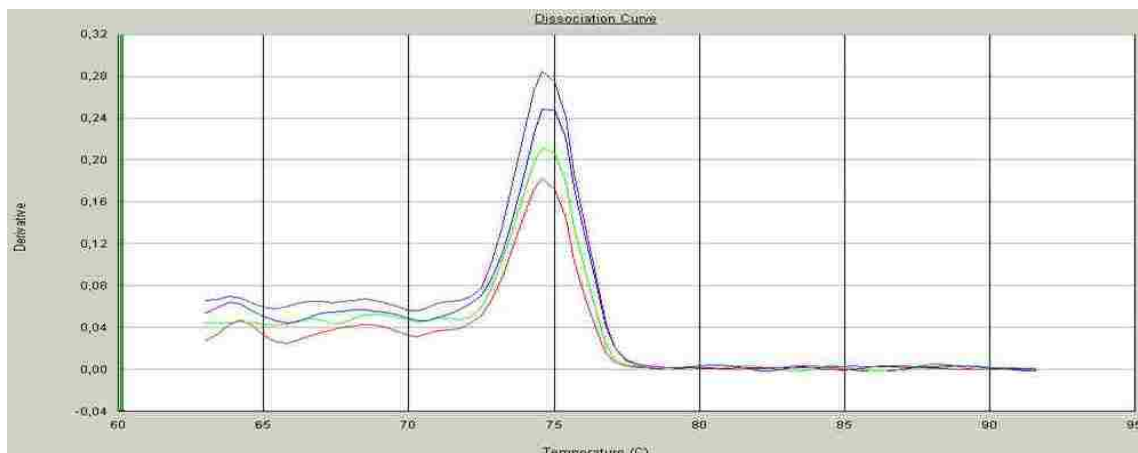
A valós idejű (real-time) PCR kísérleteinkhez az Applied Biosystems cég ABI PRISM 7000 készülékét használtuk. Ez a technika abban különbözik a hagyományos PCR-től, hogy egy fluoreszcens festék és egy kamera segítségével a szintetizálódó DNS mennyiségét a reakció során folyamatosan nyomon követhetjük.

Többféle fluoreszcens jelölés alkalmazható, mi a SYBR green festéket használtuk, amely specifikusan a duplaszálú DNS-hez kötődik, és kötődött formában erősen fluoreszkál. Mivel azonban ez a festék nem tesz különbséget a DNS szekvenciák között, a termék mellett esetlegesen képződő más duplaszálú DNS-eket – szennyeződést, a primer rossz helyre bekötődéséből származó termékeket, vagy az egymáshoz kötődő primerek

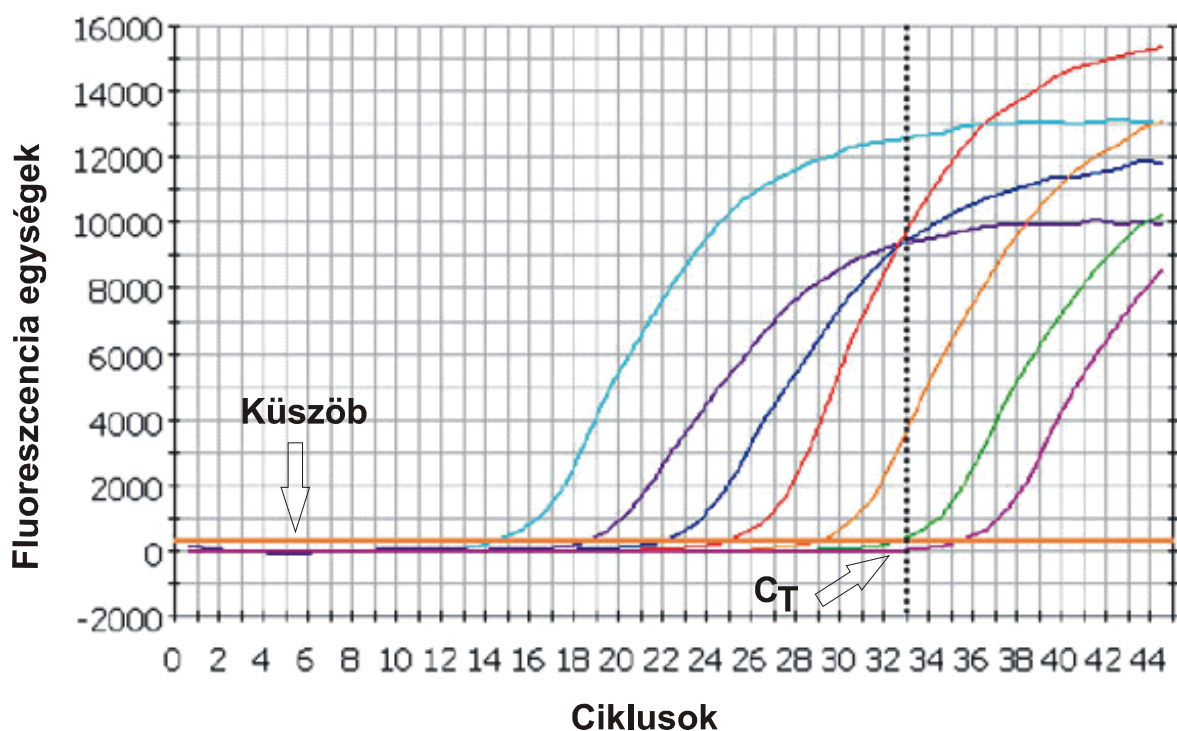
által képződő duplaszálú DNS-t – is detektálhatunk vele. Emiatt a keletkező termékeket ellenőrizni kell. Erre is többféle módszer kínálkozik, mi disszociációs görbével (4. 10. ábra) határoztuk meg, hogy egyforma DNS fragmentek szintetizálódtak-e. Ehhez a PCR reakció végén a készülék felmelegíti a mintákat, amelyekben az olvadási pontjukat elérve a kettős szálú DNS molekulák disszociálódnak. Az olvadási pont a DNS szekvenciájától függ, ezen a termékre specifikus hőmérsékleten a DNS olvadását a fluoreszcencia csökkenése jelzi.

Templátként cDNS-t használtunk, ezért az eredményekből az mRNS-szintről kaptunk információt, ezáltal pedig az adott gén kifejeződésének mértékére következtethettünk. Minél nagyobb mértékű tehát az adott gén kifejeződése, annál több PCR-terméket várhatunk.

96 csövet tartalmazó mintatartókat használtunk, ami egyszerre sok minta vizsgálatát teszi lehetővé. A reakciókhoz az Applied Biosystems cég SYBR GREEN PCR Master Mixét, 0,5-0,5 µl 10 µM-os primert és 6 µl cDNS-t mértünk össze. Minden cDNS mintát három ismétlésben vittünk fel, és a belőlük kapott értékeket a kiértékelés során átlagoltuk, ami a pipettázási hibák kiküszöbölése miatt fontos. Ahogy a keletkező DNS mennyisége nő a reakció során, egyre több festék tud kötődni hozzá, és egyre nő a fluoreszcencia. Ezt a kamera minden mintánál külön detektálja. A fluoreszcencia egységeket a készülék a ciklusszám függvényében ábrázolja (4. 11. ábra). A kiértékelés szempontjából a reakció lineáris része érdekel bennünket, itt húzzuk meg a küszöbértéket (az ábrán narancssárgával jelölve). Ez a határ a görbék alján található, közel az alapvonalhoz. Azt a pontot, ahol a fluoreszcencia metszi a küszöbértéket, azaz azt a ciklusszámot, ami a küszöbfluoreszcencia eléréséhez kell, C_T -nek nevezzük. Minél hígabb a minta, annál későbbi ciklusszámnál éri el a fluoreszcenciája a küszöböt. Ily módon a „real-time” PCR relatív kvantitálásra is alkalmas, az adott gén expressziójának szintjéről kapunk pontos információt. Ehhez azonban egy belső kontroll használata is szükséges, amelyről feltételezhető, hogy a kísérleti körülmények között nem változik az expressziója. Erre a célra búza 26S rRNS szekvenciája alapján tervezett primereket használtuk. Az eredményeket a CT módszert és a geNorm statisztikai analízist alkalmazó REST program csomaggal értékeltük és ábráztuk (Pfaffl 2001).



4. 10. ábra Disszociációs görbe. A real-time PCR reakció végén a készülék felmelegíti a mintákat, amelyek az olvadási pontjuktól függő hőmérsékleten disszociálnak, és ezt a kamera a fluoreszcencia megváltozásaként érzékeli. Az X-tengelyen a hőmérséklet, az Y-tengelyen pedig a fluoreszcencia egységek negatív deriváltja van feltüntetve.



4. 11. ábra. Amplifikációs diagram. Ahogy nő a PCR termékek mennyisége, a hozzá kötődő SYBR green mennyiségével arányosan nő a fluoreszcencia, amit a készülék a ciklusszám függvényében ábrázol. Ezen kell meghatároznunk a küszöbértéket a lineáris amplifikáció tartományában. Azt a pontot, ahol a fluoreszcencia eléri a küszöböt, C_T -nek nevezzük.

Az általunk alkalmazott kísérleti paraméterek a következők voltak:

50-100 ng cDNS (6 µl)

PCR primerek (olvadási hőmérséklet 60-61 °C) 5 pmol

8 µl PCR reakció keverék (ABI, SYBR-green PCR Master Mix)

Reakciótérfogat: 14 µl

Reakció körülmények: 94 °C 10' 1 ciklus; 94 °C 15'', 60 °C 1' 45 ciklus

A reakciókhoz ABI PRISM 7000 készüléket, az eredmények kiértékeléséhez a hozzátartozó szoftvert használtuk. A valós idejű PCR kísérletekben az alábbi primerpárokat használtuk:

Ta26S (26S rRNS)

Forward: 5' TTGTTATGCCAATAGCA 3'

Reverse: 5' GCGGCGCAGCAGTTCT 3'

EPS30

Forward: 5' GGC GTA TTT CAA GCT GCA AGA 3'

Reverse: 5' CGC CAA GCT GCG TTT CAA 3'

TFB3

Forward: 5' TATGTATGGCTGCTGTCAGG 3'

Reverse: 5' TTACCATACGACTTGCAATCAC 3'

AP2

Forward: 5' TGTGCTTCGTTTGTAGTTCTG 3'

Reverse: 5' CACAAGAAACATCGAAACAGAG 3'

EF2

Forward: 5' TCACCGTTCTGAATTTGCTG 3'

Reverse: 5' CAGACTGATAAGTTAGTTGACACC 3'

MAPKK

Forward: 5' GATTTCGGTGGCTATCTTCC 3'

Reverse: 5' TACTGTGTTTCTGACGACCC 3'

5. Eredmények

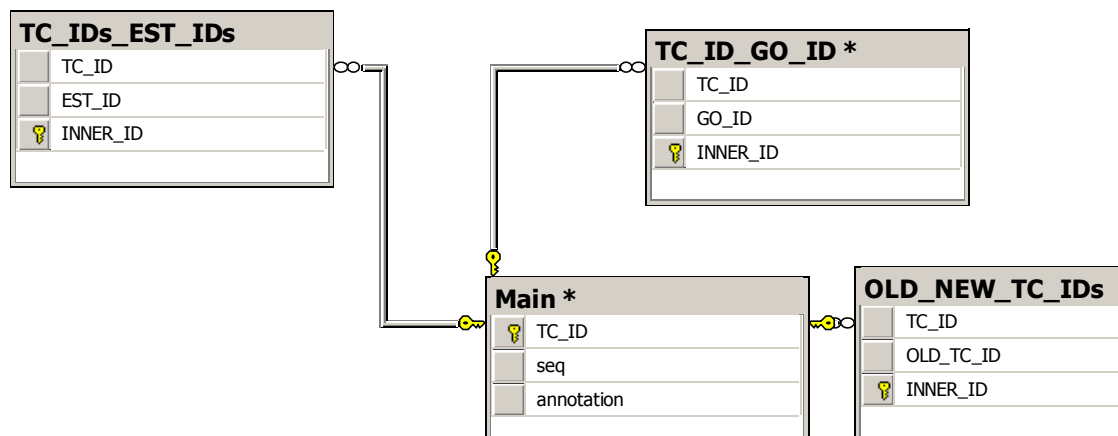
5. 1. Saját készítésű microarray kifejlesztése és tesztelése

5. 1. 1. A búza adatbázis felépítéséhez szükséges adatok

Mivel a búza genomjának szekvenciája, annak mérete és bonyolultsága miatt nem ismert, így génjeinek azonosítása, csoportosítása nem egyszerű feladat. Bár gazdasági jelentősége hatalmas, bonyolult genetikája (hexaploid volta), nem tették lehetővé, hogy modell organismusként a kutatások fő alanya legyen. Míg a rizs esetében (2007. júl. 30.-ikai adatok) alapján 213601 cDNS (core nucleotide, azaz mRNS-ként szekvenált, többször ellenőrzött, vizsgált) szekvencia van a NCBI adatbázisában, addig búza esetén, ez mindössze 8625. Természetesen a rizs genomja nem tartalmaz ekkora mennyiségű gént. Ennek a magyarázata az, hogy számos gént egymástól független több kutató csoport is meghatározta a gén bázissorrendjét.

Az EST-k száma jelentősen nem különbözik. Búzánál 1065015, rizsnél 1220709 szekvencia adat van. Jelenleg 3 olyan jelentős szolgáltató van a világon, amely az EST adatok rendszerezésével, azok contig-okba, feltételezett mRNS-ekbe, való összeszerelésével foglalkozik. Ezek a következők: NCBI (National Center of Biotechnology, <http://www.ncbi.nlm.nih.gov/>), TIGR (The Institute for Genomic Research, <http://www.tigr.org/>) és a KEGG (Kyoto Encyclopedia of Genes and Genomes, <http://www.genome.jp/kegg/>). A TIGR adatbázis ezeket a contig-okat TC-knek (tentative consensus, feltételezet konszenzus) szekvenciáknak nevezi, az NCBI Unigene-nek, míg a KEGG egyszerűen contig-nak. Ezek az adatbázisok a szekvenciák összerendezéséhez egymástól eltérő algoritmusokat használnak, ezért közöttük egy az egyhez való megfeleltetést nem lehet létrehozni. Mi alap adatbázisnak a TIGR által készített búza „Gene Index” 10 számú verzióját használtuk. Az eredmények értékelése során használtuk a „Gene Ontolgy”, gén ontológia (http://wiki.geneontology.org/index.php/GO_FAQ) rendszerét is. A rendszer RDF-XML típusú adatleíró nyelv. Kifejlesztésekkor arra törekedtek, hogy a különböző adatbázisok között kapcsolatot lehessen létesíteni, és azt hatékonyan lekérdezni.


A TIGR honlapjáról számos adat elérhető egyszerűsített letöltéssel (batch fájlként). A letöltött adatokból relációs adatbázist készítettem. Az adatbázis sémája a következő volt (5. 1. ábra).



5. 1. ábra A TIGR „Gene Index 10” adatainak az adatbázis sémája

Az elsődleges kulcsok a "Main" adattáblában a "TC_ID" mező, a többi esetben, mivel a "TC_ID" mező nem egyedi, egy saját belső automatikusan generálódó, az integritást, biztosító mezőt (INNER_ID) terveztem (a képen a kulcs szimbolumok jelzik az elsődleges kulcsokat). A gyors adatkeresés érdekében a legtöbb mezőt indexeltem. Az indexelésről azt kell tudni, hogy lelassítja az adatbevitelt, és csak akkor gyorsít jelentősen a lekérdezések sebességén, ha egyszerre csak az adatbázis kis részére vagyunk kíváncsiak. A kivételek a "Main" adattábla "seq" mezője, ahol a TC-k szekvenciáját tároltam, valamint az "OLD_NEW_TC_IDs" tábla mezői, amire nem terveztünk keresésüket. A "TC_ID" mező minden esetben a "Tentative consensus" vagy EST egyedi azonosítója (maximum 20 karakter változó hosszúságú szöveg". Hasonló típusú a "TC_IDs_EST_IDs.EST_ID" és a "TC_ID_GO_ID.GO_ID" mező. A "Main.annotation" maximum 1000 karaktert tartalmazhat, míg a Main.seq ntext típusú (bármekkora hosszúságú szöveget tartalmazhat) Az "INNER_ID" mező minden esetben egy "integer" (-2,147,483,648-tól +2,147,483,648 egész számot tartalmazó szám) típusú változó. A képen látható vonalak jelzik, hogy az adattáblák közötti kapcsolat melyik mezőkön keresztül valósul meg (elsődleges kulcs, idegen kulcs). Sajnálatos módon nem volt lehetőség, hogy „batch” fájlként könnyedén letölthessük, hogy melyik EST milyen klón

könyvtárból származik. Ez számunkra elengedhetetlen volt, hogy meghatározzuk azon TC-eket, amelyek főleg a magban expresszálnak. Ezért kerülő megoldásként egyedileg letöltöttem a 44954 TC általános összefoglalóját, ami html formátumban tartalmazta többek között a TC nevét, szekvenciáját, az őt felépítő EST-k azonosítóját és hogy melyik könyvtárból származnak. Erre a feladatra, egy egyszerű letöltés vezérlő programot írtam, ami „http_GET” metódussal működik, majd a letöltött fájlokból szövegpárosítás („string parsing” módszerrel kiválogattam a számunkra fontos adatokat. Egyedileg átnéztem a könyvtár leírásokat és kigyűjtöttem azoknak a nevét, amelyek maghoz köthetők (a kulcsszavak voltak grain, endospremium, zygote, egg cell, kernel stb.). Stratégiai döntésként az EST-kel nem, csak a TC-kel foglalkoztunk. A "Main" adattábla szerkezetét módosítottam (5. 2. ábra).

Main *			
	Column Name	Data Type	Allow Nulls
	TC_ID	nvarchar(20)	<input type="checkbox"/>
	seq	ntext	<input type="checkbox"/>
	annotation	nvarchar(1000)	<input checked="" type="checkbox"/>
	nof_EST	int	<input checked="" type="checkbox"/>
	NOF_intrested_EST	int	<input checked="" type="checkbox"/>
	PO_intrested_EST	float	<input checked="" type="checkbox"/>
			<input type="checkbox"/>

5. 2. ábra. A „Main” adattábla végső szerkezete

A "nof_EST" mező tartalmazza az adott TC-t felépítő EST-k számát, a "NOF_intrested_EST" mező tartalmazza a számunkra érdekes TC-t felépítő EST-k számát, míg a "PO_intrested_EST" mező ezek arányát tartalmazza. Az összes új mező indexelt. A TIGR „Gene Index 10” adatbázisa webes felületen össze volt kapcsolva a KEGG hierarchikus metabolikus adatbázisával. Ez úgy néz ki, mint egy Windows Explorer ablak, ahol a mappáknak 3 szintje van, ami után következik a fájllista (részletes felépítés: <http://www.genome.jp/kegg/pathway.html>). Ezeket az úgyszintén html fájlokat is letöltöttem és úgyszintén szöveg párosítás módszerével kivontam belőle az adatokat. Az adatbázishoz új táblát adtam és feltöltöttem az adatokkal (5.3. ábra). A „TC_ID” mező a

gének egyedi azonosítóját, a „primary_name, secondary_name, tertiary_name” mezők pedig a KEGG útvonal nevét tartalmazzák.

TC_metabolic *			
	Column Name	Data Type	Allow Nulls
	TC_ID	nvarchar(20)	<input type="checkbox"/>
	primary_name	nvarchar(256)	<input type="checkbox"/>
	secondary_name	nvarchar(256)	<input type="checkbox"/>
	tertiary_name	nvarchar(256)	<input type="checkbox"/>
	INNER_ID	int	<input type="checkbox"/>
			<input type="checkbox"/>

5. 3. ábra A KEGG hierarchikus adatbázis reprezentációja az adatbázisunkban

5. 1. 2. Az érdekes gének kiválasztása és az oligonukleotid tervezés

A következő lépés azon gének kiválasztása volt, amelyek szerepet játszanak a magfeltöltődés folyamatában. A keretösszeg, amit oligonukleotid szintézisére fordíthattunk kb. 82250 kapcsoláshoz volt elegendő. 70 bp hosszúságú oligonukleotidok esetén ez 1178 gént, 50 bp hosszúságúak esetén 1500 különböző oligonukleotid szintézisét tette lehetővé. 3104 TC választottunk ki oligonukleotid tervezéshez, az alábbi kritériumok alapján. Legalább két EST magi könyvtárból származik és arányuk $\geq 50\%$ -nál valamint, hogy valamilyen szabályozási folyamatban részt vegyen. Ezt annotációjuk szűrésével értük el. Kiválasztottuk azokat a TC-ket, amelyeknek leírása tartalmazta az alábbi szavak valamelyikét: transcripion, factor, kinase, phoshorilase, transporter, GTP, small nuclear, MAPK, translation, inhibitor, receptor és azokat, amelyekről még nem tudunk semmit, azaz nincs semmiféle tudásunk arról, hogy mi a funkciójuk. Ha valamilyen metabolitikus úthoz köthető a TC, akkor elég 1 magi EST és arányának $\geq 25\%$ -nak kell lennie. Próbálkoztam olyan rizs homológ búza gének megtalálásával, amelyeknek a szerepe már bizonyított a magfeltöltődésben. 9 MYB típusú transzkripciósfaktor legközelebbi búza homológját azonosítottam BLAST program segítségével. Ezen gének viszont egyáltalán nem, vagy csak alig expresszálnak magban, míg a MYB család más tagjai felül vannak reprezentálva a magi könyvtárakban. Ezért a homológián alapuló

gén kiválasztást elvetettem. Array Designer 4.0 verzióját használtam (www.premierbiosoft.com/dnamicroarray/index.html) az oligonukleotidok tervezéshez. Sense oligonukleotidokat terveztem, egy TC-re maximum tizet. A paraméterek a következők voltak: GC tartalom 30-70 % között, a olvadási pont 75 és 85°C között, a hossz 50 és 70 bp között. Az adatokat exportáltam szövegfájlba, majd az adatbázisba importáltam, az alábbi sémával (5. 4. ábra).


OLIGO *			
	Column Name	Data Type	Allow Nulls
	Accession_number	nvarchar(20)	<input type="checkbox"/>
	TC_ID	nvarchar(20)	<input type="checkbox"/>
	Sequence_length	int	<input type="checkbox"/>
	Quality	nvarchar(50)	<input type="checkbox"/>
	Rating	nvarchar(10)	<input type="checkbox"/>
	Sequence	nvarchar(70)	<input type="checkbox"/>
	Position	int	<input type="checkbox"/>
	Length	int	<input type="checkbox"/>
	Tm	float	<input type="checkbox"/>
	[GC%]	float	<input type="checkbox"/>
	[Hairpin dG(Internal)]	float	<input type="checkbox"/>
	[Hairpin Bond(Internal)]	float	<input type="checkbox"/>
	[Self Dimer dG(Internal)]	float	<input type="checkbox"/>
	[Self Dimer Bond(Inter...]	float	<input type="checkbox"/>
	[Run/Repeat Length]	int	<input type="checkbox"/>
	INNER_ID	bigint	<input type="checkbox"/>
			<input type="checkbox"/>

5. 4. ábra Az Array Designer program által tervezett oligonukleotidok tulajdonságait leíró adattábla szerkezete.

A 27569 tervezett oligonukleotid közül 23747 minősége volt jó, vagy kitűnő és a továbbiakban csak ezekkel dolgoztam. A 23747 oligonukleotid szekvenciáját FASTA formátumba exportáltam a programomból, ahol az INNER_ID mező jelenti az egyedi azonosítót, hogy azokat a BLASTN (<ftp://ftp.ncbi.nih.gov/blast/>) program segítségével analizáljam (5. 5. ábra). A BLAST programcsalád ingyenesen letölthető a NCBI honlapjáról. A program beállításai a következők voltak:

- program = BLASTN (nukleotid-nukleotid összehasonlítás)
- e = 0,001 (a kis mértékű homológiák, véletlenszerű egybeesések ne jelenjenek meg az eredményben)
- m = 9 (tabuláris formájú output fájl)
- d = TIGR_10 (az adatbázist, amihez a szekvenciákat hasonlítja én készítettem a „TIGR Wheat Gene Index 10”) adatainak felhasználásával.
- i = az általam tervezett oligonukleotidok szekvenciája FASTA formátumban

A software képes XML típusú BLAST fájlok importálására is. A kapott adatokat az alábbi táblába importáltam.

BLAST_Results *			
	Column Name	Data Type	Allow Nulls
	INNER_ID	bigint	<input type="checkbox"/>
	Query_ID	nvarchar(256)	<input type="checkbox"/>
	Subject_ID	nvarchar(256)	<input type="checkbox"/>
	[% identity]	float	<input type="checkbox"/>
	[alignment length]	int	<input type="checkbox"/>
	mismatches	int	<input type="checkbox"/>
	[gap openings]	int	<input type="checkbox"/>
	[q. start]	int	<input type="checkbox"/>
	[q. end]	int	<input type="checkbox"/>
	[s. start]	int	<input type="checkbox"/>
	[s. end]	int	<input type="checkbox"/>
	[e-value]	float	<input type="checkbox"/>
	[bit score]	float	<input type="checkbox"/>
	type	nvarchar(256)	<input type="checkbox"/>
			<input type="checkbox"/>

5. 5. ábra A BLAST eredmények formája az adatbázisban

A fontosabb mezők jelentése:

- INNER_ID = belső egyedi azonosító a koherencia fenntartása miatt
- Query_ID = az oligonukleotid azonosítója
- Subject_ID = a homológ szekvencia azonosítója

- % identity = a megegyező nukleotidok százaléka az alignment-be
- alignment_length = az alignment hossza
- type = a bevitelkor megadott azonosító, hogy több importált fájl esetén meg tudjuk őket különböztetni

Jónak tekintettük azokat az oligonukleotidokat, ahol a legmagasabb nem a teljes hosszúságú 100%-ig megegyező szekvencia részlet kisebb volt, mint 25 bp. Összesen 1504 TC-re sikerült egyedi oligonukleotidokat tervezni. Amelyik esetben többet is sikerült egy TC-re tervezni, mindig a legrövidebbet választottuk. Az SZBK „Sejtciklus Szabályozás és Stressz Adaptációs” csoportjával együttműködve az általunk tervezett oligonukleotidok kiegészültek az általuk tervezettel. Összesen 3242 oligonukleotid került fel a chipre, melyek közül 80 olyan génre volt tervezve, melynek árpa homológjának expressziója állandó volt árpa macroarray kísérletben.

5. 1. 3. A búza chip tesztelése

Az oligonukleotid szintézisét a Sigma-Aldrich, a chippek nyomtatását, hibridizálását az SZBK Funkcionális Genomikai csoportja végezte. Az eredmények kiértékelést én magam.

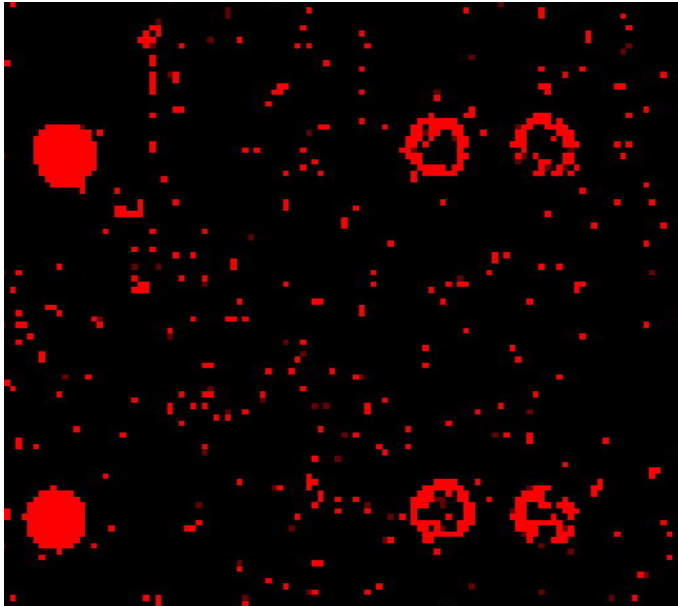
Az oligonukleotid -chippel végzett expressziós analízis lépései a következők:

- 1, Az oligonukleotid nyomtatása hordozóra
- 2, Az RNS minták cDNS-re történő átírása közben fluoreszcens festékkel történő megjelölése
- 3, A cDNS minták hibridizálása a chiphez
- 4, A nem stabil hibridizálások eltávolítása mosással
- 5, A fluoreszcenciás jelek beolvasása scanner-rel
- 6, Adat processzálas, azaz a nem megfelelő jelek eltávolítása (túl alacsony, ill. túl magas jel, inkoherencia az ismétlések között)
- 7, Az adatok normalizálása (a hibridizációs körülmények miatt bekövetkező eltérő intenzitások egy nevezőre hozása)
- 8, A biológiai ismétlések statisztikai kiértékelése, a jó és rossz pontok megjelölése

9, A jó mérések eredményeinek összehasonlítása (a tényleges kísérlet)

10, Az eredmények igazolása más módszerrel

Sajnálatos módon az első tesztelések nagyon rossz eredményeket adtak (5. 6. ábra). Minden chipen minden egyes oligonukleotid 4-szer volt felcsöppentve. Az ismétléseket statisztikai módszerrel értékeltem ki. A vizsgált gének mindösszesen a 20%-nál volt kisebb a relatív szórás (a szórás és az átlag hányadosa) mint 1, azaz az átlagos eltérés az átlagtól kisebb, mint az átlag. A hibridizálási képen a pontok nem körök, hanem gyűrűk voltak sok esetben. Az sem segített természetesen, ha az átlagtól legtávolabb eső (1 vagy 2) értéket kiszűrtem.

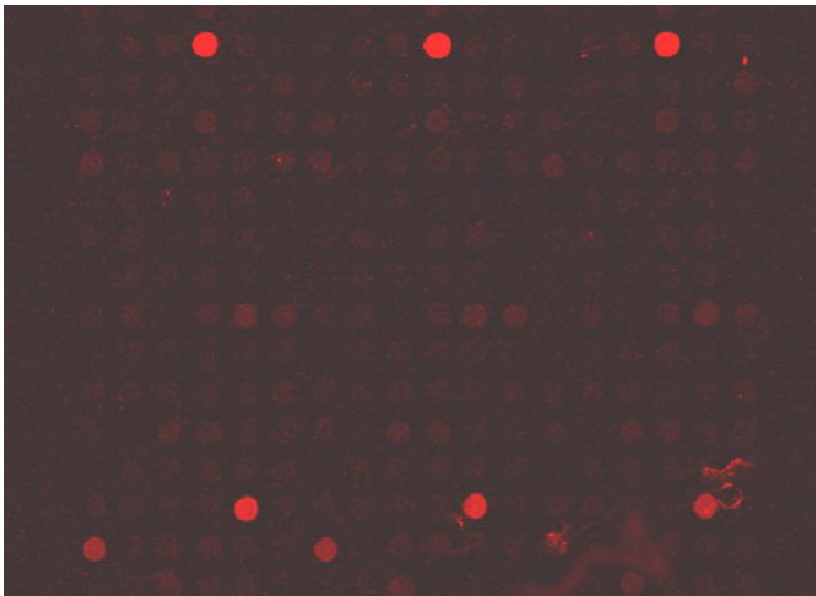


5. 6. ábra Saját készítésű oligonukleotid chip tesztelése

Az optimalizációt az SZBK Funkcionális Genomikai csoportja végezte, ezért ezt a témát nem részletezem, csak a fő pontokat emelem ki.

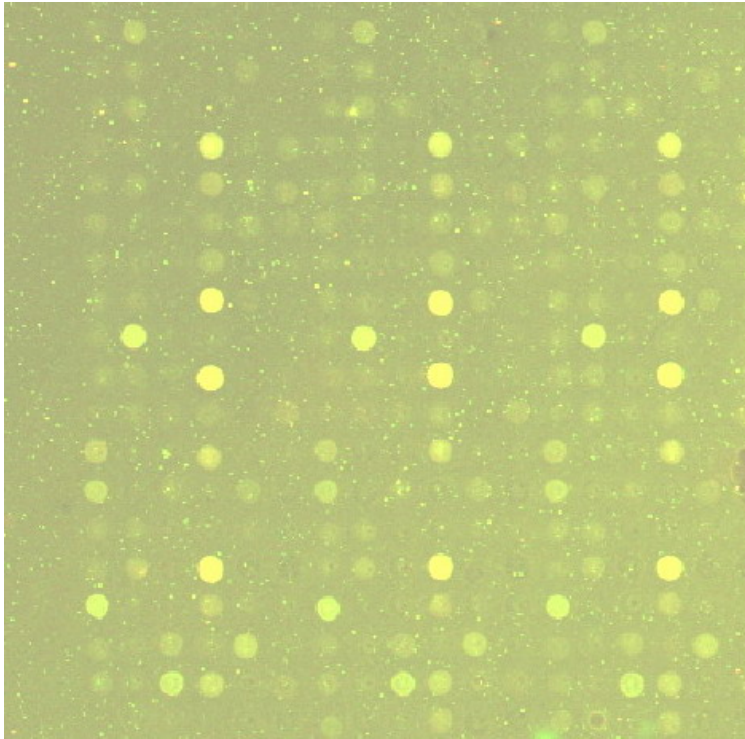
- 1, a nyomtatás (új tű)
- 2, a hordozó anyag megváltoztatása
- 3, a hibridizálási és mosási hőmérséklet, a hibridizáló és mosó oldatok optimális megválasztása
- 4, a jelölési rendszer

Az eddigi optimalizációkkal sikerült kiküszöbölni a nyomtatási problémákat és a chipen belüli ismétlések közötti eltérés már megfelel a sztenderdeknek, de sajnos a legtöbb pont jelének az erőssége olyan kicsi, hogy értékelhetetlenné teszi a felvitt oligonukleotid kb. 85%-át. Ezen jelek intenzitása alig, vagy egyáltalán nem haladja meg a háttér átlagos intenzitását (5. 7. ábra).



5. 7. ábra Saját készítésű oligonukleotid chip tesztelése I.

Ha a jelölési rendszert változtatták, egy érzékenyebb módszerre (nem teljes RNS jelölés, hanem a mRNS-eket amplifikálták fel specifikusan) akkor a háttér nagyon nagymértékben megnőtt (5. 8. ábra).



5. 8. ábra Saját készítésű oligonukleotid chip tesztelése II.

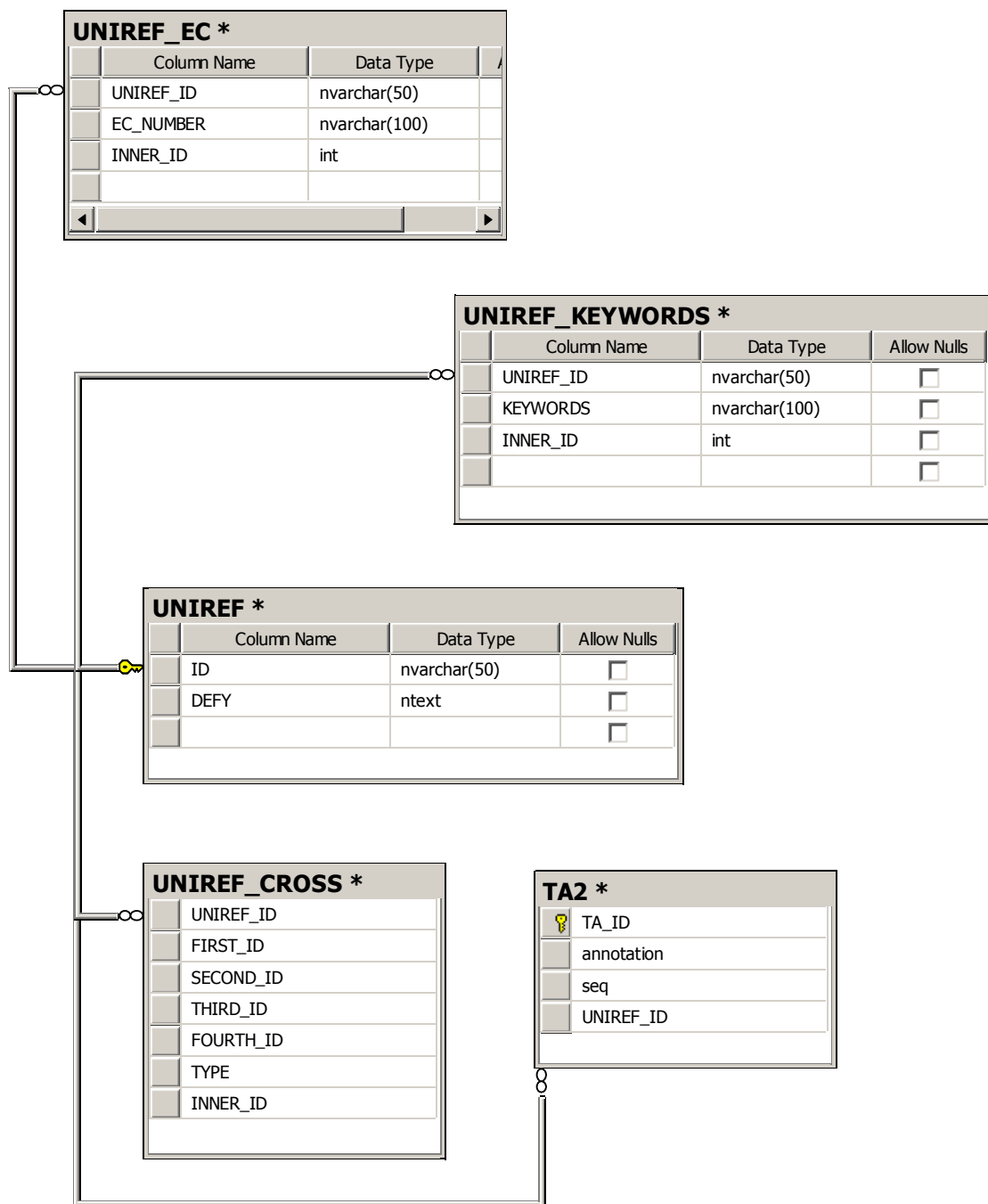
Konklúzióként levonható, hogy jelen pillanatban a rendszer még használható. A hibridizációs problémák megoldását követően azonban alkalmas lehet nagyszámú próba gyors, akár átfogóbb vizsgálatok előtti, előszűrésére. A nagy nemzetközi cégek egyre olcsóbban, egyre nagyobb sűrűségű oligonukleotid-chipeket gyártanak, már a gabonafélék, így a búza esetében is. Az Agilent cég egyedi technikája révén (a chipen magán szintetizálja az oligonukleotidokat) olyan szabadon változtatható rendszert kínál, amellyel csak egy tökéletesen tervezett, nagyon olcsó rendszer lehet versenyképes. Időközben jelentkezett egy a microarray-eket a jövőben akár teljesen kiváltó technológia is, amilyen pl. az Illumina cég rendszere, amely jelen pillanatban 4 millió transzkript szekvenálására képes egy időben, igaz jelenleg csak 25 bp hosszúságban, ami viszont általában elegendő a gének azonosítására, feltéve, ha a genom szekvenciája ismert. Ennek pontossága és érzékenysége akár lényegesen felülmúlhatja a DNS microarray-ekét.

5. 2. 1. Agilent microarray chip tervezés hő, ill. hő és szárazság stressz a búzaszem fejlődésére gyakorolt együttes hatásának vizsgálatára

A martonvásári Mezőgazdasági Kutatóintézetben Barnabás Beáta vezetésével Jäger Katalin végezte a növénynevelést és a minta előkészítést. Vizsgálataikban a szárazságstresszel szemben toleráns kansasi nemesítésű Plainsman-V és a szárazság érzékeny Cappelle Desprez francia őszi búzafajtát használták. A növénynevelés a kontroll körülményeket biztosító és a stressz-kezeléseket lehetővé tevő PGV36 és PGR15 fitotroni növénynevelő kamrákban történt. A virágzásig (antézis) kontrol körülmények (24/14 oC, napi 150 ml víz) között nevelt növényeket vízmegvonásnak vagy vízmegvonásnak és hőstressznek (34/24 oC) tették ki öt napig a szemfejlődés legkorábbi (0-5 nappal az antézis után) illetve a szemfeltöltődés korai (5-9 nap antézis után) szakaszában. A szemkezdeményeket párhuzamosa nevel kontrol és stresszelt növényekből izolálták. Az ezekből származó RNS-t használjuk majd a microarray kísérletekhez próbaként.

5. 2. 1. 1. Az oligonukleotidok tervezéshez és az eredmények kiértékeléshez szükséges adatbázis elkészítése

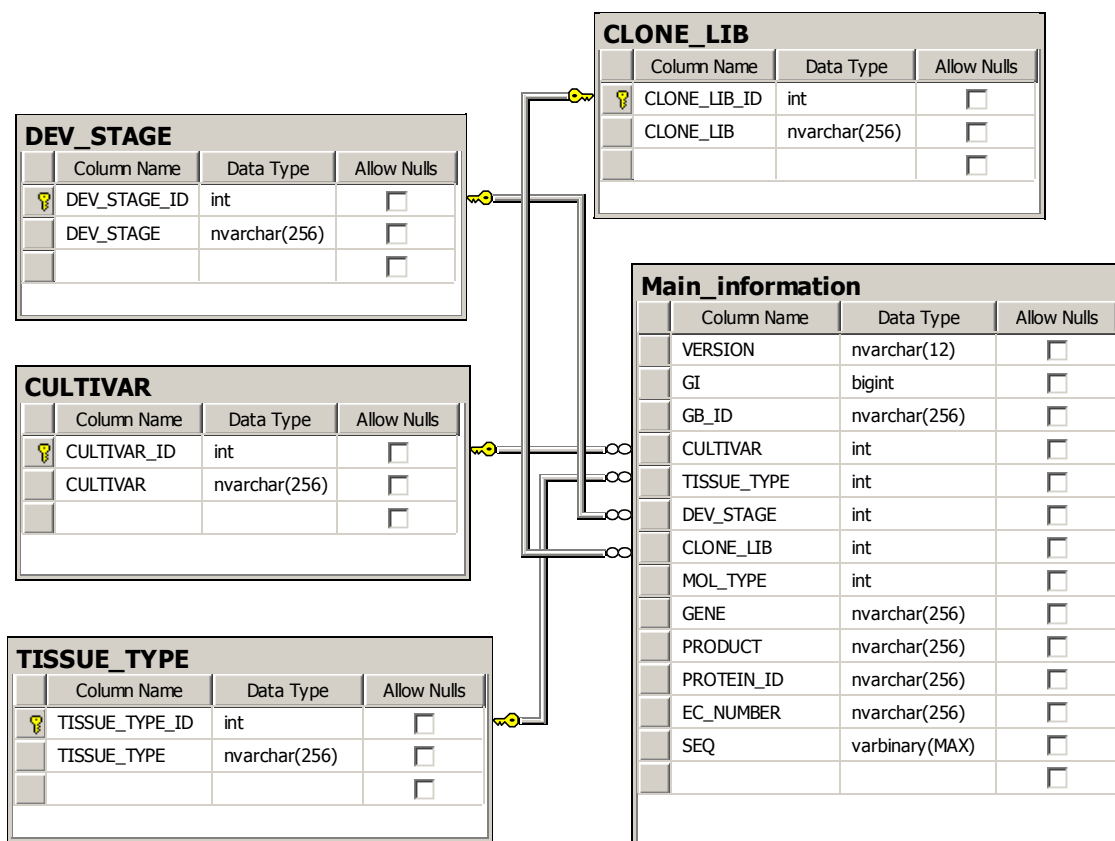
A TIGR a „Gene Index” sorozat frissítését megszüntette és egy új összeszerelő, és annotációs módszert vezetett be, a „Plant Transcript Assemblies”-t. Ennek a legújabb búza verziójának szekvencia adatait használtuk az oligonukleotidok tervezéséhez. A „batch” fájlként elérhető adatok mindösszesen az egyedi azonosító, a szekvencia és az annotáció volt. Az annotáció alapját képező UniRef adatok nem voltak elérhetőek, ahogy a TA-kat (Transcript assembly, összeszerelt, mesterséges szekvencia) felépítő EST-k adatai se. Némi fenyegetésre (5 gépről másodpercenként indítva a lekérdezéseket, megbombázzuk a szerverüket, amíg le nem töltjük a 62121 TA-t és a 257828 singleton-t (azok az EST-k, melyeket, nem sikerült contig-ba szervezni)) ezekhez is sikerült hozzájutni. Mivel a gének annotációja az UniRef adatbázison alapul, ezért az adatbázisunkban összekötöttem őket az alábbi séma szerint (5. 9. ábra).



5. 9. ábra Az UniRef adatbázis szerkezete

A TA2 adattábla tartalmazza a szekvenciát, annak egyedi azonosítóját, az annotációt és az annotációhoz tartozó UniRef azonosítót. A UniRef_EC táblába az adott UniRef azonosítóhoz tartozó EC (Enzyme catalog) számát tároltam. A UniRef_Keywords táblába az adott UniRef azonosítóhoz tartozó kulcsszavakat tároltam. Ezekre a gyors keresés miatt

volt szükség. A Uniref_Cross táblába a más adatbázisokhoz (kb. 100 darab) való elérhetőségeket tároltam, hogy egy kattintással a software beépített Web böngészőjén keresztül el lehessen érni a kívánt információt. Az UniRef adattábla tartalmazza az egyedi azonosítót és a hozzá tartozó leírást mindenféle módosítás nélkül. Ahhoz, hogy meghatározzam, hogy melyek azok a gének, amelyek a magban expresszálnak, tudnom kellett, hogy a TA-kat felépítő EST-k milyen könyvtárból származnak. Ehhez letöltöttem az NCBI-nál található összes búza EST leírását GenBank formátumban (2,2 Gb). A fájlból kiszűrtem a számunkra fontos információkat és az adatbázisba importáltam. Ezek voltak az EST GenBank „accession”-je, a GI száma, a hozzátartozó EC szám, GO számok, a szövet, a fejlődési állapot és a fajta (5. 10. ábra). Mivel itt hatalmas mennyiségű adatot kellett feldolgozni és az adatbázisba vinni, ezért a programomat nagymértékben kellett optimalizálni (IO fájlműveletek, memória kihasználás és processzor kihasználtság között kellett megtalálni a legcélravezetőbb megoldást).



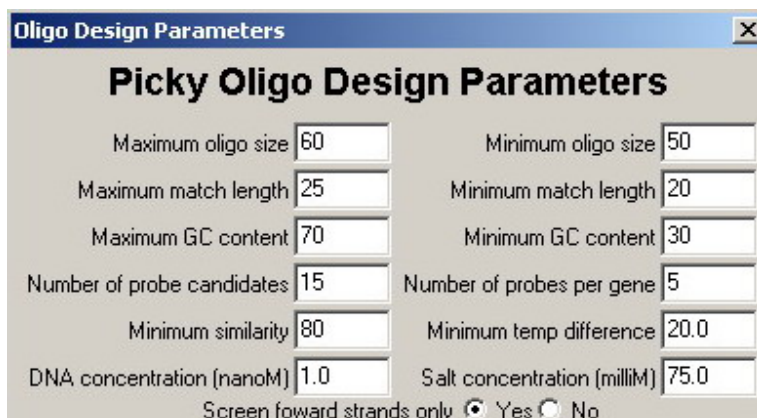
5. 10. ábra A búza EST-k adatainak tárolási sémája az adatbázisban

A fejlődési állapotot, a szövet típusát, a fajta nevét és a klónkönyvtár nevét azért tároltam külön táblába, mert az adatokba rengeteg ismétlés van. Egy integer típusú változó tárolása jóval kevesebb helyet igényel, mintha egy 256 karakterből álló szöveget mentenék el. A szekvenciákat nem szöveggént, hanem bináris formátumba alakítottam és így tároltam. Az átalakítás lényege, hogy amíg a rendszer egy karaktert 2 byte-on tárol, addig az átalakított forma mindössze 3 bit-et foglal el. Összesen 441 klónkönyvtár szerepel az adatbázisunkba. Ezek közül 156 könyvtárat tekintettük szemfeltöltődéshez köthetőnek (ld. függelék 10. 3.). Meghatároztuk a TA-k EST-inek összetételét és arányát, amit az adatbázisba mentettem el (5. 11. ábra).

Seed_related_all *			
	Column Name	Data Type	Allow Nulls
	[NOF EST]	int	<input type="checkbox"/>
	TA_ID	nvarchar(13)	<input type="checkbox"/>
	NOF_Selected_EST	int	<input checked="" type="checkbox"/>
	percentSeed_related_all	float	<input checked="" type="checkbox"/>
			<input type="checkbox"/>

5. 11. ábra A magban expreszálódnó TA-k kiválasztásához szükséges adattábla szerkezete


Azokat a TA-kat és singletonok-at, amelyeknek van legalább egy magi EST-je FASTA formátumba exportáltam (121343 szekvencia) két részletbe, mert az oligonukleotid tervezéshez használt PICKY (<http://www.complex.iastate.edu/download/Picky/>) program csak 65536 szekvenciát tud kezelni egyszerre. Az alábbi beállításokkal futattam a programot (5. 12. ábra).



Picky Oligo Design Parameters	
Maximum oligo size	60
Minimum oligo size	50
Maximum match length	25
Minimum match length	20
Maximum GC content	70
Minimum GC content	30
Number of probe candidates	15
Number of probes per gene	5
Minimum similarity	80
Minimum temp difference	20.0
DNA concentration (nanoM)	1.0
Salt concentration (milliM)	75.0
Screen forward strands only <input checked="" type="radio"/> Yes <input type="radio"/> No	

5. 12. ábra A PICKY program futtatási paramétere.

A program nagyon gyors, de nagyon sok memóriát igényel. A már említett Array Designer 2 nap alatt volt képes oligonukleotidokat találni a 3104 kiválasztott génre, a PICKY kevesebb, mint 1 óra alatt megtervezte a 65536 génre az oligonukleotidokat. Nem minden génre sikerült tervezni. Összesen 358621 oligonukleotidot terveztem, melyek közül 331667 volt egyedi. Ennek oka, hogy hasonló szekvenciákra, ha nem ugyanabban a csoportba voltak, ugyanolyan oligonukleotidot tervezett a software. A program további előnye, hogy ha homológiát talál 2 vagy több szekvencia között, akkor kiszámolja a maximális kereszthibridizálási hőmérsékletet, és ha az kisebb egy küszöbértéknél akkor a tervezett oligonukleotidot eldobja. Ez a tulajdonsága a programnak, alkalmas arra, hogy az olyan nagyon konzervált, de nem azonos szekvencia részletek, mint például az ATP kötésért felelős régió a kinázokban, nem célpontjai az oligonukleotid tervezésnek. Ezen kívül képes több hasonló szekvenciára egy közös oligonukleotidot tervezni. A kapott oligonukleotidok adatait az alábbi adattáblába töltöttem föl (5. 13. ábra).

PICKY			
	Column Name	Data Type	Allow Nulls
	[PROBE DESIGN]	nvarchar(100)	<input type="checkbox"/>
	LENGTH	int	<input type="checkbox"/>
	TEMP	float	<input type="checkbox"/>
	NONTARGET	nvarchar(256)	<input type="checkbox"/>
	TYPE	char(1)	<input type="checkbox"/>
	SITE	int	<input type="checkbox"/>
	TEMP_BAD	float	<input type="checkbox"/>
	 ID	bigint	<input type="checkbox"/>
	is_good	bit	<input checked="" type="checkbox"/>
	nof_similar	int	<input checked="" type="checkbox"/>
	HAS_KO	bit	<input checked="" type="checkbox"/>
	nof_EST	int	<input checked="" type="checkbox"/>
	nof_selected_EST	int	<input checked="" type="checkbox"/>
	percent_of_EST	float	<input checked="" type="checkbox"/>
	duplicate	nvarchar(50)	<input checked="" type="checkbox"/>
			<input type="checkbox"/>

5. 13. ábra Az oligonukleotidok jellemzésére használt adattábla sémája.

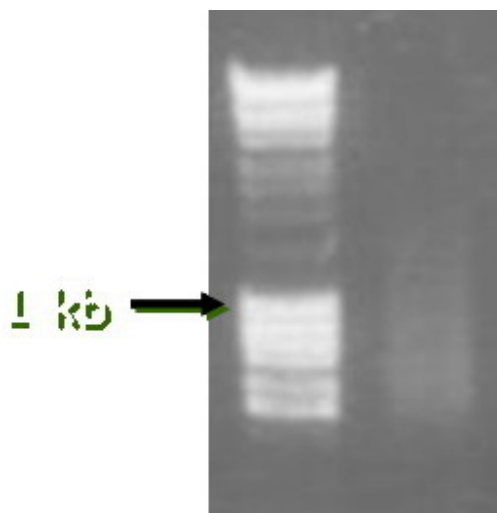
Az első 7 sor a PICKY által tervezet oligonukleotidok adatait tartalmazza, mint szekvencia, hossz, olvadási pont, a legjobban homológ szekvencia azonosítója, hogy egyedi, avagy közös oligonukleotid-e, hogy mi a kezdőpontja a szekvencián belül és , hogy mi a legközelebbi keresztreagáló szekvenciával az olvadási pont. Az „ID” mező egyedi azonosító. Az ismétlődő szekvenciákat eltávolítottam az adatbázisból, és a maradékot FASTA formátumban exportáltam ki. Ezek után a már ismertettet módon BLAST program segítségével analizáltam tovább a már említett paraméterekkel. Érdekes módon bizonyos szekvenciák BLAST-olása során, a NOD32-es vírusirtóm fertőzést állapított meg és törölte a kimeneti szöveg (txt) típusú fájlt. A vírusirtómat ki kellett kapcsolnom. Összesen 4482524 adatsort vittem fel a Blast_Results adattáblába. Itt szeretném megjegyezni, hogy a sokak által használt és szeretett EXCEL program mindösszesen 65536 adatsort képes használni, ami kb. az 1/70-eke csak a most felvitt adatoknak. Jónak fogadtam el az oligonukleotidot, ha BLAST eredménye legalább 5 darab 100%-ig homológ és az oligonukleotidok hosszával megegyező alignment TA-t vagy singleton-t talált. Akkor is jónak fogadtam el, ha a tökéletesen megegyező TA-k és

singletonokon kívül nem volt olyan keresztreagáló gén, ahol az alignment hosszának és a megegyező nukleotidok százalékának a szorzata kisebb, mint 90%. Más esetekben az oligonukleotidot rossznak tekintettem és a további szűrésekből kizártam. Az „is_good” mezőben tároltam ezeket az információkat. Összesen 174877 oligonukleotid kapott megfelelő minősítést. A „nof_similar” mezőben a 100% teljes hosszúságban homológ TA-k, singletonok számát tároltam. A következő szűrési lépés az volt, hogy azon oligonukleotidok közül, melyek pontosan ugyanazt a gént ismerik fel, csak a leghosszabbat hagytam meg, a többiekénél a „duplicate” mező értékét „Igaz”-ra állítottam be. 111087 oligonukleotid távolítottam el az eredeti listából. Egy „Agilent” chipen jelenleg maximálisan 244000 oligonukleotid helyezhető el, de lehetőség van a chipet kisebb részekre osztani. Lehetőség van 2, 4 vagy 8 párhuzamos kísérlet elvégzésére egy chipen. Mi szakmai és anyagi megfontolások után, úgy döntöttünk, hogy egy 8 osztatú chipet tervezünk, amin egyszerre kb. 15000 oligonukleotidot lehet részenként elhelyezni. A következő lépésben végigmentem az összes génen, amire oligonukleotid terveztünk. Minden egyes génhez megkerestem, hogy melyik oligonukleotid ismerik fel. Ezek közül azt az egy oligonukleotid hagytam meg, amelyik a legtöbb gént ismerte fel. Ez a lépés arra is jó volt, hogy feldúsítsam azon oligonukleotidok számát, amelyek a különböző fajták számos allélja közül a közös, megegyező szekvenciájú szakaszokrara legyenek tervezve. Összesen 35975 potenciális oligonukleotid maradt. Ezek közül azokat válogattuk ki, ahol az oligonukleotid által felismert gének EST-i közül legalább kettő a maghoz és annak fejlődéséhez köthető. A chipre így 14966 különböző oligonukleotid került. Az előző fejezetben említett kontroll génekre még további 230 oligonukleotid lett tervezve, amelyeket kiszűrtünk. Mivel volt még 244 szabad hely a chipen ezért ezeket is kiválasztottuk. A maradék 12 szabad helyre 3 oligonukleotidot még 4-4 példányba felvittük. Ezek közül egy olyan gén volt, aminek összesen 2 EST-je ismert, egy olyan, aminek 34 és egy olyan, aminek több mint 1000. Tehát egy feltehetően alacsonyan, közepesen és erősen expresszáladó gént választottunk ki kontrolnak, hogy megnézzük a chipen belül van-e eltérés a más helyeken szintetizált, de azonos szekvenciájú oligonukleotidok között. Az Agilent rendszere ezenkívül minden esetben fenntart még 536 pontot, ahol a különböző kontrollok (bakteriális gén, üres pont) helyezkednek el.

5. 3. 1. Búza petesejt és korai (8 órás) zigóta transzkripció analízise

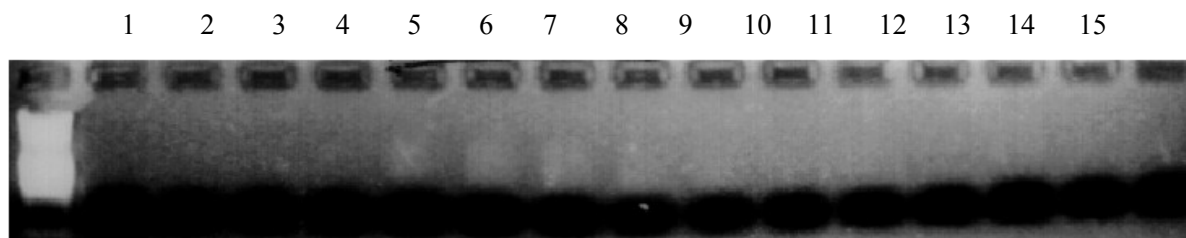
Mivel a petesejt és zigóta izolálás nagyon munka és anyag igényes, így ezek a sejtek csak korlátozott mértékben álltak rendelkezésünkre, ennek következtében a tisztított mRNS-ek mennyisége nagyon kevés volt, ezért az EST szekvenálás módszerét alkalmaztuk. Az eredményeket „Real-time PCR” technológiával ellenőriztük. Mindkét alkalmazás jellemzője, hogy nagyon kis mennyiségű RNS mintát lehet velük vizsgálni. Más génexpressziót vizsgáló módszer nem alkalmazható. A sejtek izolálását a MTA Mezőgazdasági Kutatóintézetének Sejtbiológiai Csoportjában végezték.

A cDNS könyvtár elkészítését a CLONTECH Laboratories cég SMARTtm PCR cDNA Library Construction Kit alapján végeztük. Ismétlésekként átlag 50 sejtől mRNS-t tisztítottunk, majd átírtuk cDNS-sé. A cDNS-ek számát specifikus oligók felhasználásával PCR technológiával megsokszoroztuk. Agaróz gélelektroforézissel ellenőriztük le a mRNS tisztítás, cDNS készítés és az amplifikáció hatékonyságát (5. 14. ábra).



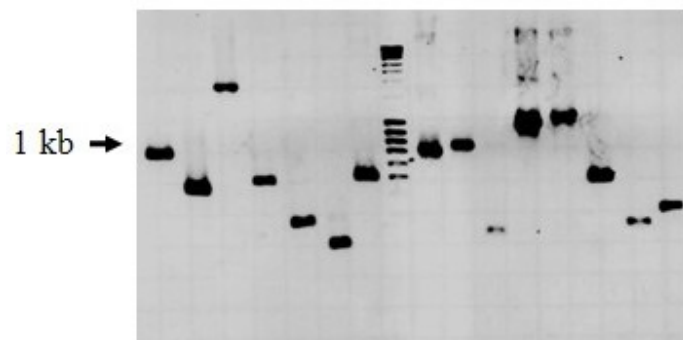
5. 14. ábra a cDNS készítés és amplifikáció ellenőrzése agaróz gélelektroforézissel

A kész cDNS-eket Sfi I restrikciós endonukleázzal megemésztettük, majd oszlopkromatográfiás elválasztás után a nagyobb méretű cDNS molekulákat elkülönítettük és velük dolgoztunk tovább (5. 15. ábra).



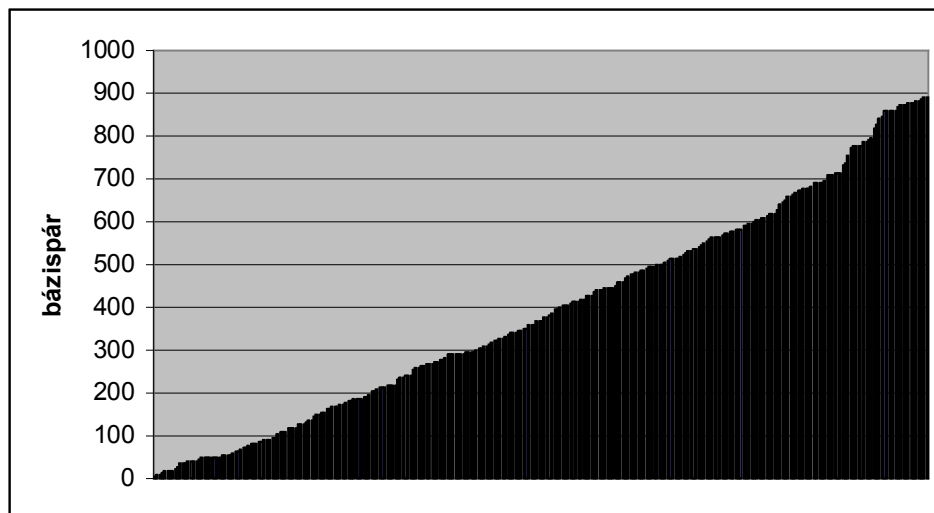
5. 15. ábra Oszlopkromatográfia utáni ellenőrzés

A DNS molekulákat vektorhoz ligáltuk, majd fág fehérjékbe csomagoltuk. □-komplementációval ellenőriztük a beépülés sikerességét. A pete könyvtár esetén ez az érték 80% fölött volt. A zigóta könyvtár esetén, hiába próbáltuk optimalizálni az eddigi lépéseket, mindösszesen kb. 60 %-os volt a cDNS inszertet hordozó fágok aránya. Meghatároztuk a független fágok számát. Pete esetén ez volt, ami igen jó érték. Sajnos a zigóta könyvtár esetében ez már nem volt ilyen jó, de elfogadható 1×10^6 / ml. A pete könyvtár esetében a fágokat megsokszoroztuk, de a zigóta könyvtár esetében, mivel annak minősége (üres fágok nagy aránya) nem volt elég jó, ezt a lépést nem végeztük el. Ezután egyedi fágokat izoláltunk, majd in vivo kivágással plazmiddá alakítottuk át, és DNS-t tisztítottunk belőlük. A vektorhoz tervezett primerek használatával, PCR technikával meghatároztuk az inzertek méret szerinti eloszlását (5. 16. Ábra). A legalább 200 bp hosszúságú klónok közül választottuk ki a megszekvenálókát. A szekvenálásokat a MacroGene



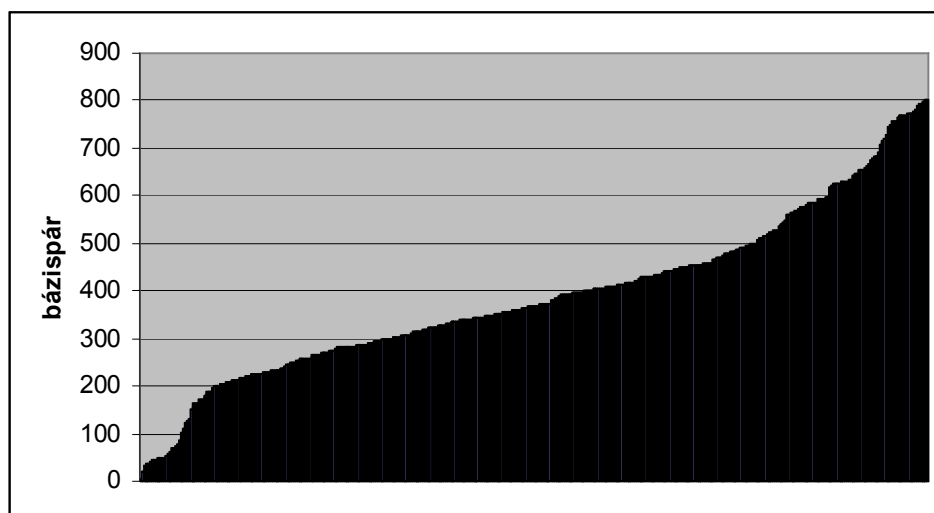
5. 16. ábra Inszertek méretének meghatározása vektor specifikus primerekkel

A pete könyvtár inzertjeinek méret szerinti megoszlása a következő volt.



5. 17. ábra A pete könyvtár inzertjeinek méret szerinti megoszlása

A zigóta könyvtár inzertjeinek méret szerinti megoszlása a következő volt.



5. 18. ábra A zigóta könyvtár inszertjeinek méret szerinti megoszlása

A kapott EST adatok feldolgozása

A kapott szekvenciákat, egy általam készített szoftverrel, megtisztítottuk az 5' ill., 3' vektor szekvenciákat tartalmazó részekről, a rossz minőségű szakaszokról valamint a polyA szekvenciáktól. A felismeréshez BLAST algoritmust alkalmaztam, amely a legalkalmasabb módszer lokális szekvencia részletek azonosságának a megállapítására. A software lehetőséget ad, hogy megadjuk az 5' ill., a 3' vektor szekvenciát, beállítsuk a polyA szakasz hosszát, valamint a rossz minőségű szekvencia felismerési paramétereit. Mi az alábbi paraméterekkel dolgoztunk. Az inszerció, delécio és a mismatch egységesen -1 pontot ért. A nukleotid egyezés +1 pontot (ezek a paraméterek is szabadon választhatók). Ha a hibák száma 3-nál kevesebb, akkor az adott szekvenciát felismertnek tekintettük. A program kimenete egy olyan szöveg típusú fájl, amit az Excel képes beolvasni). Ez a fájl tartalmazza az eredeti fájl nevét, a régi szekvenciát, az új szekvenciát, és hogy felismerte-e a vektor szekvenciákat, polyA-t és hogy az eredeti szekvencia melyik bázisánál. A program multiple FASTA (egy fájlban belül sok szekvencia) formátumban elmenti az új jó szekvenciákat. Volt néhány szekvencia, amelynek mérete nem érte el a 100 bázispárt. Ezekkel nem foglalkoztunk tovább. Az olyan szekvenciákkal, melyeknél az 5' vektor régiót nem ismerte fel a program (13 db) egyedileg megvizsgáltam. Ezeknél, a

mintáknál szekvenálási gondok voltak és a további vizsgálatokban nem foglalkoztam velük.

A	B	C	D	E	F	G	H	I	J	K	L
név	eredeti szekvencia	hossz	5' vektor	5' vektor hely	3' vektor	3' vektor hely	polyA	polyA hely	új szekvencia	hossz	
1	ZIG400-TEXGGGGTTTTNNNN	862	True	68	True	553	True	521	GGGACTTATG	453	
2	ZIG401-TEXNNCCGGGTNNNN	792	True	64	True	326	True	296	GGGCTGATCA	232	
3	ZIG402-TEXNNTTCTTTCTGGA	850	True	57	False		True	386	GAAGAGCGG	339	
4	ZIG403-TEXNTTCTCTGAGAC	847	True	54	True	307	True	269	GAGGGCGGAC	215	
5	ZIG404-TEXNNNTTCTTTCTGG	858	True	58	True	651	True	617	GGTCTTTGGT	559	
6	ZIG405-TEXGGGTCTTTCTGAG	852	True	58	False		False		GGGCCGTCG	794	
7	ZIG406-TEXGGGGTCTTTCTG	857	True	58	False		True	397	GAGTATGTGG	339	
8	ZIG407-TEXGGGGATGTTTCGG	850	True	58	True	380	True	349	GGGTGGAGG	291	
9	ZIG408-TEXNAAAGAGAGANNI	838	True	73	False		False		GGACATCCAG	765	
10	ZIG409-TEXGAGATTANCNNNN	868	True	68	True	414	True	381	GGACATGAGC	313	
11	ZIG410-TEXNTTTTCTGCGGAC	846	True	55	True	431	True	400	GATGGAGCGC	345	
12	ZIG411-TEXTTTTCTGGGAGCC	841	True	53	True	526	True	494	GGGAGAGGA	441	
13	ZIG412-TEXGGTAAGGNNNNNT	666	True	61	False		False		GCGTAGTGCT	47	
14	ZIG413-TEXGGGGNNNTTCT	857	True	59	True	350	True	317	GAGGTGCTTA	258	
15	ZIG414-TEXNNNNCTTCGGTA	851	True	56	True	358	True	327	GCTCTAAAAA	271	
16	ZIG415-TEXNNTTTCTTTCTGG	857	True	58	True	488	True	458	GGATGTCAAT	400	
17	ZIG416-TEXNGGATTTTAANNI	867	True	67	True	520	True	486	GGGAGAGGA	419	
18	ZIG417-TEXGGGGNNNTTCT	859	True	59	True	462	True	431	GGGTGCCATG	372	
19	ZIG419-TEXNAAANTTTTCTGG	859	True	58	True	463	True	430	GGAACCTATT	372	
20	ZIG422-TEXAGGTGTTTNNNNI	865	True	68	True	380	True	346	GAAGGAGGG	278	
21	ZIG423-TEXNTTTTATCTGGAA	856	True	59	True	283	True	248	GCATGACCTC	189	
22	ZIG426-TEXNNNNNNNTTCTGG	856	True	58	True	281	True	246	GGGGACCTTG	188	
23	ZIG427-TEXNAAAGGAATNNI	796	True	66	False		True	362	GGCAGAAGG	296	
24	ZIG430-TEXNTTTTCTGGA	853	True	57	True	296	True	265	GGTGTGTTT	208	
25	ZIG431-TEXNGGGTGTNNNN	853	True	65	True	445	True	414	GGGAGGAAG	349	
26	ZIG434-TEXNTTTTCTGGGA	854	True	57	True	329	True	299	GGACATCCAG	242	
27	ZIG436-TEXNAGATTTNNNA	859	True	66	False		True	501	GAGACAAACT	435	
28	ZIG438-TEXNNNTTTTCTGGAC	854	True	56	True	459	True	429	GGCTATTCTG	373	
29	ZIG439-TEXNNNTTTTCTGACG	852	True	54	True	587	True	554	GTTACCGGCA	500	
30	ZIG440-TEXNNNTTTTCTGG	850	True	58	True	485	True	450	GGACCATCTC	392	
31	ZIG441-TEXAAGAAANNTTCT	865	True	60	True	498	True	461	GAATTACACT	401	

5.18. ábra Az EST adatok megtisztításának kimeneti fájlja EXCEL formátumban megnyitva.

A megtisztított szekvenciákat BLASTN program segítségével analizáltam. A keresendő adatbázist a TIGR Plant Transcript Assemblies Wheat szekció adatainak felhasználásával készítettem el. A kapott eredményeket adatbázisba vittem, majd azokat a saját EST-ket, ahol a hozzá tartozó adatbázis szekvenciák átfedést mutattak más saját EST-nkhez tartozó adatbázis szekvenciával (átfedő szekvencia részleteknél 95 %-os azonosság volt a küszöb érték) egy FASTA típusú fájlba exportáltam (n dimenziós halmazelméleti probléma). A fájlokat CAP3 (Huang és Madan, 1999) program segítségével tovább vizsgáltam (paraméterek: legalább 80%-os egyezés, 25 bázispáros átfedés). Ez a program nem lokális hasonlóságokat keres, mint a BLAST, hanem

megpróbálja összeszerelni egy, vagy több contig-ba a bemeneti szekvenciákat. A pete szekvenálások esetén összesen 15 contig-ot szerelt össze a szoftver (5. 19. ábra).

annotáció	EST-k száma
TA64547_4565 Putative dihydrolipoamide dehydrogenase [Oryza sativa (japonica cultivar-group)]	7
TA54806_4565 T6D22.2 [Arabidopsis thaliana (Mouse-ear cress)]	4
TA58264_4565 Glyceraldehyde-3-phosphate dehydrogenase, cytosolic 3 [Zea mays (Maize)]	3
TA57874_4565 40S ribosomal protein S8 [Oryza sativa (Rice)]	3
DR739462	3
CJ567765 Putative ubiquinone oxidoreductase subunit [Oryza sativa (Rice)]	3
TA61725_4565 Ribosomal protein L18a-like [Oryza sativa (japonica cultivar-group)]	3
TA58459_4565 Glyceraldehyde-3-phosphate dehydrogenase, cytosolic [Hordeum vulgare (Barley)]	2
TA50617_4565 Putative ubiquitin-associated (UBA) protein [Oryza sativa (japonica cultivar-group)]	2
CJ616936 Putative 40S ribosomal protein S3 [Triticum aestivum (Wheat)]	2
TA62694_4565 Putative OsCTTP [Oryza sativa (japonica cultivar-group)]	2
TA62564_4565 Ribosomal protein L15 [Oryza sativa (japonica cultivar-group)]	2
TA57365_4565 GTP-binding nuclear protein Ran/TC4 [Vicia faba (Broad bean)]	2
TA59714_4565 Hypothetical protein OJ1513_F02.133 [Oryza sativa (japonica cultivar-group)]	2
TA53484_4565 40S ribosomal protein S21 [Oryza sativa (Rice)]	2

5.19. ábra A több pete EST-t tartalmazó contigok annotációja.

190 EST egyedi volt a 243 megszekvenált génből, azaz az általunk megszekvenált szekvenciák 80%-a. Ez a cDNS bank alacsony redundanciáját igazolja. A vizsgált 243 szekvencia közül 178 mutatott nagy mértékű homológiát már ismert búza génekhez (90%-nál nagyobb homológia). A maradék 65 szekvencia egyáltalán nem, vagy csak elenyésző mértékű szekvencia egyezést mutatott (50 bp maximális egyezés és 50 % nukleotid egyezés), azaz ezek új, eddig búzában még nem ismert géneknek tekinthetők. Ezeket a

szekvenciákat tovább elemeztük. Megvizsgáltuk, hogy az átírt fehérje szekvenciához van-e hasonló, más növényekben előforduló már ismert fehérje, vagy más növényekben írtak-e le hasonló géneket (BlastX, BlastN programok használata, beállítások, taxid: 3193, csak növény; adatbázis: est, nem-redundáns mRNS, nem redundáns fehérje szekvencia) . 39 esetben találtunk növényi homológot, míg 26 esetben nem (az „e” értéknek 0,01-nél kisebbnek kellett lennie). 17 EST-nkhez tartozó gént írtak már le búza pete könyvtárból, ezek közül egyet (melyet EPS30-nak neveztünk el) részletesebben is vizsgáltunk

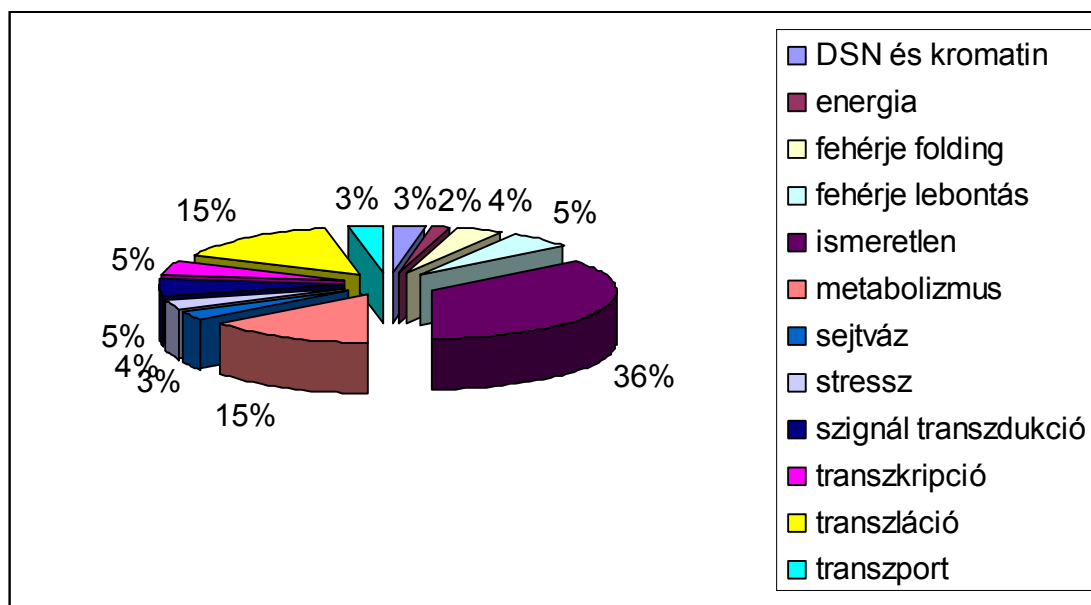
A 8 órás zigóta könyvtár szekvenciáit hasonló módon vizsgáltam (5. 20. ábra). Itt összesen 451 jó minőségű szekvenciával dolgoztam. Azért használtam több mintát, mert búza petesejtből már készítettek EST könyvtárat, de fiatal még nem osztódó zigótából nem. Osztódó 2 napos zigóta búza EST könyvtár úgyszintén létezik.

A zigóta szekvenciák esetében az egyedi EST-k aránya 73% volt. Összesen 69 zigóta EST-nkhez volt már ismert zigótából (2 napos) izolált gén, tehát itt az átfedés a két adatbázis között nagyobb volt. Az átfedéseket figyelembe véve ez összesen 35 gént jelentett.

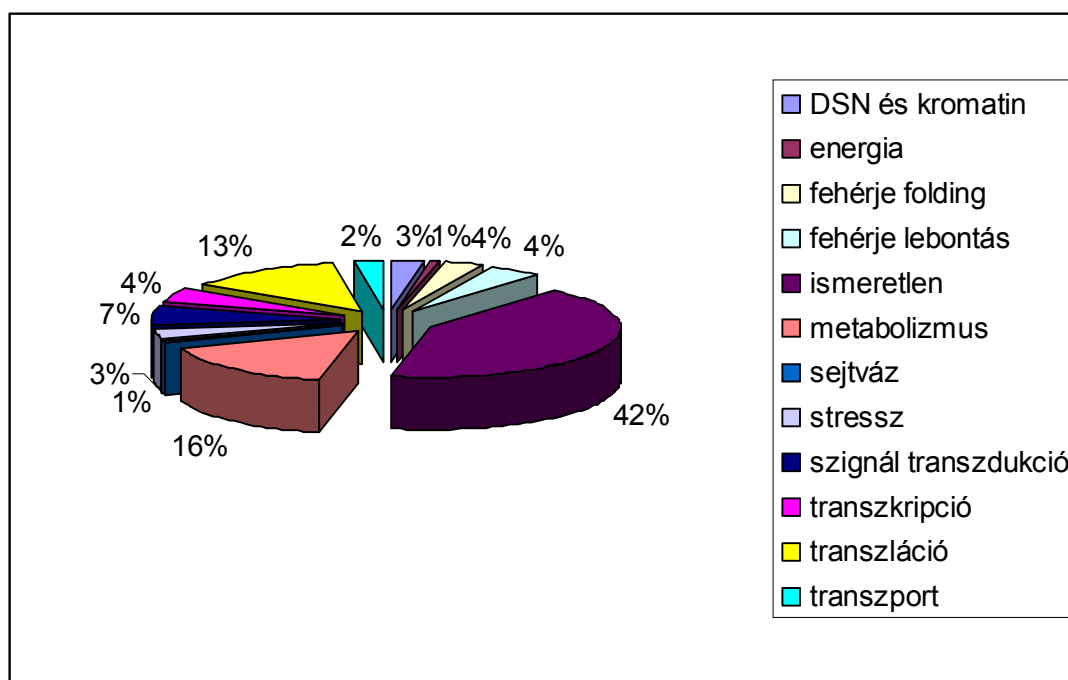
annotáció	EST-k száma
CD927088 Lactoylglutathione lyase [Oryza sativa (Rice)]	13
CJ590407 Cytosolic heat shock protein 90 [Hordeum vulgare (Barley)]	7
CK194926 60s ribosomal protein L21 [Triticum aestivum (Wheat)]	6
CK217687 OJ000223_09.15 protein [Oryza sativa (japonica cultivar-group)]	6
TA92871_4565 SelT/selW/selH selenoprotein domain, putative [Oryza sativa (japonica cultivar-group)]	5
TA53131_4565 ATP synthase D chain, mitochondrial [Arabidopsis thaliana (Mouse-ear cress)]	5
TA53398_4565 Ribosomal protein l34 [Triticum aestivum (Wheat)]	4
CJ570397 Disease-resistant-related protein [Oryza sativa (Rice)]	3
TA63098_4565	3
CA642500	3
CK168245 60S ribosomal protein L10-3 [Oryza sativa (Rice)]	3
DN949012 40S ribosomal protein S28 [Kluyveromyces lactis (Yeast)]	3
CJ687469 Ribosomal protein L15 [Oryza sativa (japonica cultivar-group)]	3
CJ658480 60S ribosomal protein L44-like [Arabidopsis thaliana (Mouse-ear cress)]	3
TA63131_4565 Hypothetical protein P0413G02.21 [Oryza sativa	3

(japonica cultivar-group)]	
TA55927_4565 Transport protein subunit-like [Arabidopsis thaliana (Mouse-ear cress)]	3
TA74680_4565 Proteasome inhibitor-like protein [Oryza sativa (japonica cultivar-group)]	2
TA93774_4565 Hypothetical protein OJ1014_C08.10 [Oryza sativa (japonica cultivar-group)]	2
TA93140_4565	2
TA59204_4565 26S proteasome regulatory particle non-ATPase subunit12 [Oryza sativa (japonica cultivar-group)]	2
CJ559219 Putative RNA-binding protein 10 [Oryza sativa (japonica cultivar-group)]	2
CJ569624 Peptidyl-prolyl cis-trans isomerase [Oryza sativa (japonica cultivar-group)]	2
CK210751 MATE efflux family protein, putative [Oryza sativa (japonica cultivar-group)]	2
CJ627838	2
CJ605250 Putative zinc finger transcription factor ZFP33 [Oryza sativa (japonica cultivar-group)]	2
TA60745_4565 Expressed protein [Oryza sativa (japonica cultivar-group)]	2
TA62095_4565 Putative oxoglutarate/malate translocator protein [Arabidopsis thaliana (Mouse-ear cress)]	2
TA61940_4565 Putative 60S ribosomal protein L12 [Oryza sativa (japonica cultivar-group)]	2
TA71217_4565	2
CJ656975 F10A5.19 [Arabidopsis thaliana (Mouse-ear cress)]	2
CK206972	2
TA60297_4565 Putative polynucleotide adenylyltransferase [Oryza sativa (indica cultivar-group)]	2
CK199029	2
CK196444 40S ribosomal protein S8 [Oryza sativa (Rice)]	2
TA59346_4565 Putative elongation factor 1 beta [Hordeum vulgare (Barley)]	2
TA56989_4565 Eukaryotic translation initiation factor 5A1 [Triticum aestivum (Wheat)]	2
CJ565245 Putative organic solute transporter [Oryza sativa (Rice)]	2
TA52896_4565 Ribosomal protein L35A [Zea mays (Maize)]	2
CJ618691 Hypothetical protein P0036D10.5 [Oryza sativa (japonica cultivar-group)]	2
DR741188 Putative immunophilin [Hordeum vulgare (Barley)]	2
5. 20. ábra A több zigóta EST-t tartalmazó contigok annotációja	

A szakirodalom felhasználásával minden egyes klónt feltételezett funkciója alapján besoroltam valamilyen kategóriába (5. 21. ábra, 5. 22. ábra).



5. 21. ábra A pete EST-k funkcionális csoportosítása szakirodalom alapján



5. 22. ábra A zigóta EST-k funkcionális csoportosítása szakirodalom alapján.

Az általunk kapott szekvenciákat kiegészítettük más könyvtárak (2 napos zigóta, ill. pete) szekvenciáival is. A 2 új könyvtár EST-jén is elvégeztem a BLAST analízist. A

programom segítségével automatizált bioinformatikai elemzést végeztem. A folyamat a következő lépésekből áll.

a, Adatbázisomból lekérdeztem az EST-khez tartozó hozzá homológ (legalább 90% nukleotid sorrend megegyezés és legalább 100 bp alignment hossz (mindkét paraméter változtatható)) TA azonosítót

b, Adatbázisomból lekérdeztem a TA-khoz tartozó „KO” azonosítót vagy „GO” azonosítót, olyan módon, hogy egy EST-hez ugyanaz a „KO”, ill. „GO” azonosító csak egyszer tartozhat.

c, Adatbázisomból lekérdeztem a „KO” ill. „GO” azonosítóhoz tartozó adatokat.

d, Az előzetesen elkészített „KEGG”, ill. „GO” fanézethez (treeview, windows explorer szerű nézet) párosítottam az adatokat.

e, A program végigment a fanézet összes csomópontján (node) és az összes hozzá tartozó bejegyzést (gént, oligonukleotid, entitást), megszámlolta.

f, Azokat a funkciókat, annotációkat, melyekhez egyik vizsgált halmazban sem található referencia, láthatatlanná tette, mert különben átláthatatlan adathalmazt kapnánk („GO” annotációból 120000 bejegyzés van).

Mivel az SQL nyelv lehetővé teszi a kapcsolt lekérdezéseket (ez gyorsabb, mint a ciklusos lekérdezés, ahol az első eredményhalmaz eredményeiből készítünk egy új lekérdezést és egy új eredményhalmazt, majd ezt folytatjuk addig, amíg el nem jutunk az utolsó eredményhalmazig), ezért az első három pontot egy utasítással oldottam meg. Ez a lekérdezés SQL nyelven így néz ki a KEGG adatbázishoz való összevetésben.

```
SELECT          DISTINCT      BLAST_RESULTS.QUERY_ID,
KEGG_HIERARCHY.ID,          KEGG_HIERARCHY.PARENT_ID,
KEGG_HIERARCHY.GROUP_ID1,   KEGG_HIERARCHY.GROUP_ID2,
KEGG_HIERARCHY.GROUP_ID3,   KEGG_HIERARCHY.GROUP_ID4,
KEGG_HIERARCHY.PATH, KEGG_HIERARCHY.PATH_NAME
FROM            TA_ETAE_KO INNER JOIN          KEGG_HIERARCHY ON
TA_ETAE_KO.KO = KEGG_HIERARCHY.PATH
INNER JOIN      BLAST_RESULTS  ON  TA_ETAE_KO.TA_ID  =
BLAST_RESULTS.[SUBJECT ID]
```

WHERE BLAST_RESULTS.[TYPE] = 'ZIGOTA' AND
 BLAST_RESULTS.[ALIGNMENT LENGTH] >= 100 AND
 BLAST_RESULTS.[% IDENTITY] >=90

Jelen esetben, minden egyes előbb említett könyvtárhoz (ez persze nem kötelezően egy könyvtárhoz tartozó szekvenciák, lehet teljesen más alapján kiválasztott gének egy csoportja is). A program futása során csak azokkal a szekvenciákkal számol, amiről már van részletes ismeretünk. Az eredményt számos formátumban (xls , txt, pdf, rtf, html) képes elmenteni és újra betölteni. Az ábrákon kiragadott részletek láthatók, mert a teljes elemzés több mint 40 oldal (5. 22. ábra, 5. 23. ábra).

PATH_NAME	2d zigóta	egg cell	pete	8h zigóta	Picky
Metabolism	128	118	34	75	1682
Carbohydrate Metabolism	39	27	11	31	460
Glycolysis / Gluconeogenesis	16	6	5	2	77
hexokinase	0	0	0	0	9
glucose-6-phosphate isomerase	0	0	0	0	1
fructose-1,6-bisphosphatase I	0	0	0	0	2
fructose-bisphosphate aldolase, class I	1	4	0	0	4
triosephosphate isomerase (TIM)	1	0	1	1	3
glyceraldehyde 3-phosphate dehydrogenase	7	1	2	1	33
phosphoglycerate kinase	0	0	1	0	1
phosphoglycerate mutase	0	0	0	0	2
enolase	3	0	1	0	0
pyruvate kinase	2	0	0	0	3
pyruvate dehydrogenase E1 component, beta subunit	0	0	0	0	2
dihydrolipoamide dehydrogenase	0	0	0	0	2
pyruvate decarboxylase	0	1	0	0	4
alcohol dehydrogenase (NADP+)	0	0	0	0	4
acetyl-CoA synthetase	0	0	0	0	1
aldose 1-epimerase	0	0	0	0	6
phosphoglucomutase	2	0	0	0	0
Citrate cycle (TCA cycle)	4	0	2	4	27
citrate synthase	0	0	0	0	4
ATP citrate (pro-S)-lyase	0	0	0	0	1
aconitate hydratase	2	0	0	0	0
aconitate hydratase 1	0	0	0	0	1
isocitrate dehydrogenase	0	0	0	0	1
2-oxoglutarate dehydrogenase E1 component	0	0	1	0	1

5. 22. ábra Pillanatkép a szoftverből. A különböző könyvtárakból származó EST-k és az általunk tervezett oligonukleotidok száma a KEGG adatbázis funkcionális kategóriáiban.

Name	2d zigóta	egg cell	pete	8h zigóta	Picky
molecular function	1542	1060	472	837	366262
structural molecule activity		2			131
structural constituent of cuticle		2			131
structural constituent of chitin-based cuticle		2			131
structural constituent of adult chitin-based cuticle		2			131
transporter activity	44	39	19	26	27305
protein transporter activity	2		2	5	1095
arginine targeting transporter activity			1		202
protein translocase activity					3
protein-transporting ATPase activity			1	5	47
bacteriocin-transporting ATPase activity			1	5	47
protein carrier activity					5
colicin transporter activity	2				833
group A colicin transporter activity	2				833
alcohol transporter activity					14
polyol transporter activity					14
glycerol transporter activity					12
glycerol-3-phosphate transporter activity					12
inorganic phosphate antiporter activity					12
galactitol transporter activity					2
galactitol permease activity					2
permease activity					149
efflux permease activity					149
formate efflux permease activity					149
d transporter activity	2	2		1	1542
dopamine transporter activity	1	1		1	1101
sodium symporter activity	1	1		1	1101

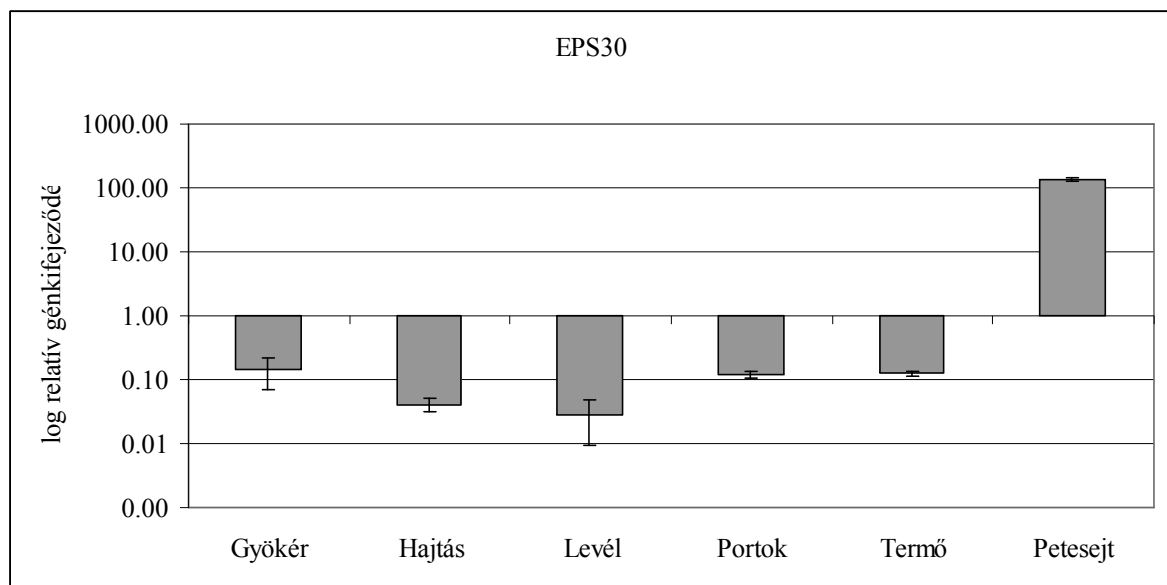
5. 23. ábra Pillanatkép a szoftverből. A különböző könyvtárakból származó EST-k és az általunk tervezett oligonukleotidok száma a GO adatbázis kategóriáiban.

5. 3. 3. Pete és zigóta specifikus gének meghatározása RT-PCR technológiával

Mivel célunk egyebek mellett petesejt-specifikus gének azonsítása volt, ezért kiválasztottunk egy a petesejtben igen erőteljes kifejeződést mutató (13 EST szekvencia került már be az adatbankokba, mind petesejt könyvtárakból) specifikus gént. Ennek jele

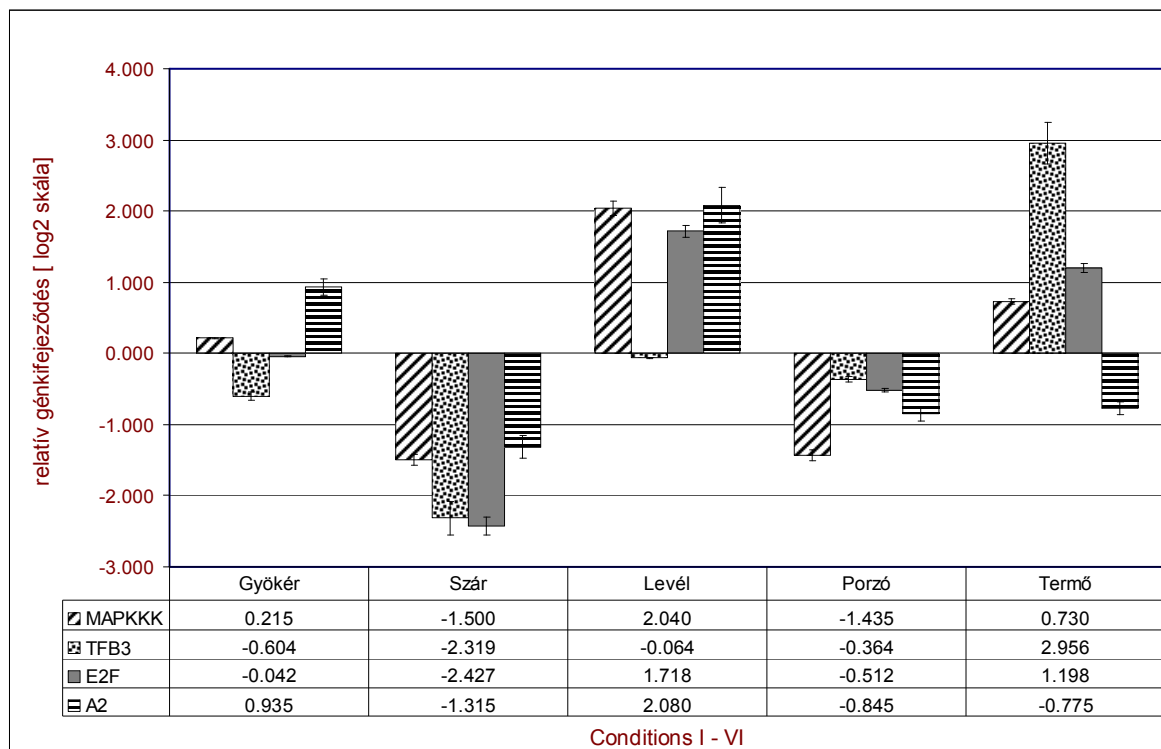
EPS30 (érett petesejt specifikus 30) és egy ubiquitin/SUMO-szerű kis fehérjét kódol, amelynek a fehérje degradációban van szerepe.

A gén specifikus kifejeződését real-time kvantitatív RT-PCR (RT-QPCR) segítségével ellenőriztük, amihez RNS-t izoláltunk különböző búza szervekből és petesejtből. Ez utóbbi esetekben a kis sejtszám miatt, a cDNS populációt SMART PCR (Clontech, „SMART PCR cDNA synthesis protokoll”) segítségével lineárisan megsokszoroztuk (15 ciklus). Kimutattuk, hogy a gén kifejeződése a pete sejtekben nagyságrendekkel meghaladja az egyéb szervekben mérhető értéket (5. 24. ábra).



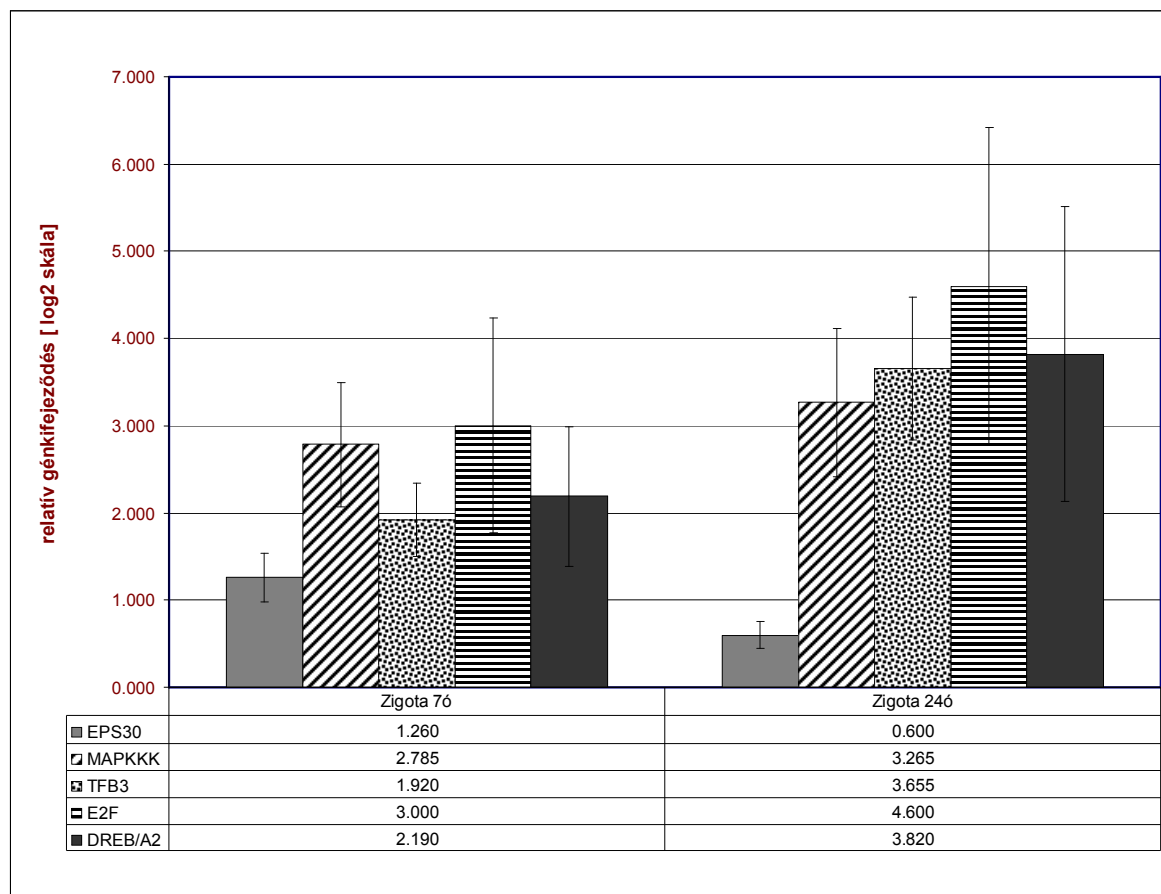
5. 24. ábra AZ EPS30 gén relatív génkifejeződése búzában. A génkifejeződést a Ta26S rRNS gén kifejeződésére normalizáltuk, és az EPS30 génnek a vizsgált mintákban mért átlagos kifejeződéséhez (=1) viszonyítva ábrázoltuk (három ismételtes átlaga és a szórás).

A zigóta cDNS bank EST szekvenciái között több olyan potenciális gén szerepelt, amelyek részt vehetnek a megtermékenyítést követő folyamatok elindításában. Ezek között három transzkripciós faktor (TFB3, A2-like és E2F) és egy MAPKK-kináz fehérje génje tűnt a legígéretesebbnek. Ezek relatív expressziós mintázatát is meghatároztuk a különböző növényi szervekben (5. 25. ábra). A gének több növényi szervben is kifejeződnek, elsősorban levélben és gyökérben, és a DREB/A2 transzkripciós faktor kivételével a termőben is magas relatív génkifejeződést mutatnak.



5. 25. ábra A zigóta cDNS könyvtárból azonosított néhány szabályozó fehérjét kódoló gén relatív kifejeződése növényi szervekben. A génkifejeződést a Ta26S rRNS gén kifejeződésére normalizáltuk, és minden egyes gént az adott gén vizsgált mintákban mért átlagos kifejeződéséhez (=1) viszonyítva ábrázoljuk (három ismétlés átlaga és a standard deviáció).

A vizsgált gének szerepét a megtermékenyítés folyamatában úgy vizsgáltuk, hogy a petesejtben, illetve a megtermékenyítést követően a zigótában (7 óra illetve 24 óra) összehasonlítottuk a relatív génkifejeződést (5. 26. ábra).



5. 26. ábra A zigóta cDNS könyvtárból azonosított néhány szabályozó fehérjét kódoló gén illetve a petesejtben erőteljes kifejeződést mutató EPS30 gén relatív kifejeződése petesejtben a megtermékenyítés előtt és után. A génkifejeződést a Ta26S rRNS gén kifejeződésére normalizáltuk, és minden egyes gént az adott gén petesejtben mért átlagos kifejeződéséhez (=1) viszonyítva ábrázoltuk (három ismétlés átlaga és a szórás).

A vizsgált gének közül az EPS30 gén a megtermékenyítést követően csökkenő, a többi gén növekvő kifejeződést mutatott, ami összhangba van azonosításuk módjával. A későbbiek folyamán kerülhet sor arra, hogy a teljes hosszúságú cDNS-eiket izoláljuk és petesejtekbe mikroinjektálva a megtermékenyítés mesterséges elindítására való képességüket értékeljük.

6. Eredmények megvitatása

6. 1. A növényi petesejt transzkripciós profilja

A növények életciklusa egy diploid sporofitikus és egy haploid gametofitikus nemzedék váltakozásaként zajlik. Jelenlegi ismereteink nagy része a domináns sporofitikus életszakaszra vonatkozik, annál is inkább, mert a növényi gametofiton csak néhány sejtből áll és azok is a sporofita szövetekbe mélyen beágyazva helyezkednek el. Napjaink modern kutatási eszközei azonban lehetővé teszik, hogy a növényeknek azt a rejtett életszakaszát is alaposabban megvizsgáljuk. T-DNS inszerciós mutagenesis segítségével lúdfűben több, mint száz olyan mutációt azonosítottak, amelyek a gametofiton fejlődését érintik (Pagnussat és mtsai. 2005). DNS mikroarray (DNS-chip) segítségével feltérképezték a gametofiton génkifejeződési mintázatát és 225 gametofiton specifikus gént azonosítottak (Yu és mtsai. 2005).

A petesejt speciális szerepet tölt be a gametofitonon belül, hiszen a megtermékenyítést követően belőle fejlődik az új sporofiton nemzedék. Nem tudjuk azonban, hogy ebben a képességében mennyi az anyai hatás és mennyi a megtermékenyítés szerepe. Arról is kevés információ áll rendelkezésünkre, hogy a növényi petesejt mennyire aktív és milyen folyamatok zajlanak benne. Mindezekről átfogó képet nyújthat a petesejt génkifejeződési mintázatának vizsgálata és annak összehasonlítása más növényi sejtekével.

A lúdfű petesejt vizsgálata ilyen szempontból nem lehetséges, kis mérete és izolálásának nehézkessége miatt. A kukorica és a búza esetében azonban már évek óta rendelkezésünkre állnak azok a technikák, amelyek lehetővé teszik a petesejtek rutinszerű izolálását. Ezeknek a technikáknak a kidolgozásában Barnabás Beáta kutatócsoportjának úttörő szerepe volt. Erre a tapasztalatra és technikai lehetőségre alapozva, közös kutatások keretében, elkészítettünk egy búza petesejt cDNS könyvtárat és megvizsgáltuk, hogy milyen gének fejeződnek ki ebben a speciális sejttípusban. Mint kiderült, ez a megközelítés nem csak bennünk merült fel, és Sprunck és munkatársai 2005-ben leközölték 404 búza petesejtben expresszáldó gén azonosítását és elsődleges jellemzését.

Ezek a cDNS-ek a szekvencia adatbázisokban elérhetők, így összehasonlíthatóak voltak a laboratóriumunkban azonosított 251 búza petesejt cDNS szekvenciával. Az azonos szekvenciák száma nagyon kevés volt, így 237 petesejtben expresszáldó új gént azonosítottunk. Ezek közül 73 még teljesen ismeretlen búza szekvencia, míg 188 korábban már más cDNS könyvtárakból előkerült. Mindkét cDNS szekvencia készlet arra enged következtetni, hogy a petesejtben jelentős anyagcsere aktivitás és intenzív fehérjeszintű változások zajlanak, azaz a növényi petesejt koránt sem annyira nyugvó, mint azt korábban feltételezték (Russell 1993).

Kumlehn és munkatársai (Kumlehn és mtsai. 2001) szintén beszámoltak már búza petesejtből, mégpedig a parthenogenetikus fejlődésre képes „Salmon” genotípus petesejtjeiből, készített cDNS bankról. A mintegy ötszáz azonosított cDNS szekvencia részletet azonban nem tették nyilvánossá. Azóta még több petesejt cDNS (EST) szekvenciát azonosítottak, de ezek sem kerültek még be a nyilvános szekvencia adatbázisokba. Elmondhatjuk azonban, hogy a három laboratórium összességében mintegy ezer-ezerötszáz különböző búza petesejtben expresszáldó gént azonosított, ezeknek az átfogó összehasonlító vizsgálata azonban még várat magára.

A fent említett búza petesejt cDNS könyvtárak kapcsán még meg kell említenünk, hogy egyedül az általunk készített könyvtár készült phagemid vektorban és olyan módon, hogy a petesejtben expresszáldó gének legalább felét ($1,5 \times 10^6$ gént) reprezentálja. Az amplifikált cDNS könyvtár szinte korlátlan mennyiségben áll rendelkezésre és lehetővé teszi hagyományos molekuláris szűrési technikák alkalmazását. A másik két cDNS könyvtár lényegesen kisebb számú, néhány ezer, plazmid DNS-be klónozott cDNS fragmentumot tartalmaz.

6. 2. A zigóta és a kétsejtes embrió transzkripció profilja

Sprunck és munkatársai (Sprunck és mtsai. 2005) kétsejtes búza embrióból is készítettek cDNS könyvtárat és ennek felhasználásával 289 EST szekvenciát határoztak meg. A petesejt és a kétsejtes zigóta transzkript profiljának összehasonlítása azt jelezte, hogy az embrióban intenzívebb transzkripció és transzláció folyik és természetesen a sejtszétváláshoz kapcsolódó gének kifejeződése is kimutatható. Kukorica kétsejtes

embriókból is azonosítottak cDNS szekvenciákat (Okamoto és mtsai. 2005), mégpedig ebben az esetben a bazális és apikális sejtből külön-külön, és megállapították, hogy a két sejt között is jelentős génkifejeződési különbség van.

Ezek a vizsgálatok azonban nem nyújtottak információt arról, hogy mikor következik be jelentős változás a génaktivításban a petesejt megtermékenyítését követően. Ennek a kérdésnek a megválaszolásában segíthet az általunk a megtermékenyítést követően hét órával begyűjtött zigótákból készített cDNS könyvtár, és az ebből származó cDNS-ek szekvenálása. Az eddig azonosított 369 zigóta EST szekvencia alapján azt mondhatjuk, hogy a megtermékenyítést követő hetedik órában már jelátvitelben résztvevő fehérjék és több transzkripciós faktor is megjelenik, ami arra utal, hogy a zigóta saját génkészlete már aktiválódott. Ennek az egyértelmű igazolása azonban még további kísérleteket igényel.

Búza, és általában növényi, zigóta cDNS könyvtárral, tudomásunk szerint csak a mi laboratóriumunk rendelkezik. Az általunk azonosított 369 zigóta szekvencia közül 81 eddig ismeretlen búza gént vagy allélt reprezentál. A 288 már ismert szekvencia közül közül 23 a petesejt, 22 pedig a kétsejtes embrió cDNS könyvtárban is előfordul. Természetesen a rendelkezésre álló szekvencia mennyiség nem elégséges átfogó összehasonlító elemzés elvégzéséhez. További szekvenálásra és egyedi gének kifejeződésének vizsgálatára van szükség ahhoz, hogy a petesejt megtermékenyítését követő transzkripciós változásokról valós képet kaphassunk.

6. 3. Genomikai megközelítési lehetőségek a búza kutatásában

A búza (*Triticum aestivum* L.) az egyik legfontosabb gazdasági növény, amelynek kutatása kiemelt jelentőséggel bír. Ugyanakkor a modern genomikai kutatási módszerek alkalmazása a búza esetében nem könnyű feladat. Ezt a problémát a búza genom hatalmas mérete okozza. A termesztett búza poliploid növény, amely hatalmas méretű kromoszómákkal rendelkezik. Az allohexaploid genom összességében mintegy 16 gigabázisnyi (Gb) DNS-t tartalmaz (Bennett és Smith 1976), szemben a rizs 450 Mb-nyi genom méretével (Sasaki és Sederoff 2003). A különbségért elsősorban repetitív, nem kódoló, szekvenciák felelősek, a két faj génkészlete ugyanis nagymértékben hasonló (pl.

(Sorrells és mtsai. 2003). A rizs genom becslések szerint valamivel kevesebb, mint 40000 fehérje kódoló gént tartalmaz (Bennetzen és mtsai. 2004). A búza gének számára vonatkozó becslések elég tág intervallumba szórnak, de minden bizonnyal jóval meghaladják a rizs gének számát.

A rizs, és az árpa mint a gabonafélék kutatásának modell növényei, sok hasznos információval szolgálhatnak a búza kutatása és akár nemesítése számára is (Appels és mtsai. 2003, Feuillet és Keller 2002, Ware és mtsai. 2002). Ezeknek az információknak az alkalmazása és a búza specifikus tulajdonságainak kutatása megköveteli azonban, hogy a búza esetében is rendelkezésre álljanak az alapvető genomikai megközelítési lehetőségek. Mivel a búza genom mérete nem teszi lehetővé a teljes genom szekvencia meghatározását az elkövetkezendő néhány éven belül, ezért előtérbe került a búza cDNS szekvenciák (EST-k) minél nagyobb léptékű összegyűjtése az adatbankokban. Ehhez változatos, a búza különböző fejlődési stádiumait és eltérő környezeti hatásokra adott válaszait reprezentáló cDNS bankok előállítására van szükség. Laboratóriumunk a búza petesejt és zigóta cDNS bank elkészítésével ehhez a törekvéshez járult/járul hozzá. Az általunk meghatározott 620 EST szekvencia közül 154 az, ami az adatbázisokban fellelhető, nagyságrendileg több százezer, búza EST szekvenciával nem mutatott azonosságot (95% alatti a DNS szintű azonosság), azaz feltehetően eddig ismeretlen búza gént vagy allélt reprezentál. Ezen kívül számos esetben olyan ismert gének szekvenciáját is sikerült bővítenünk, amelyeket már reprezentált az általunk meghatározott szekvenciával átfedő adatbázis szekvencia. Azokban az esetekben pedig, ahol az általunk meghatározott szekvencia már ismert volt, új információként kerülhetett az adatbázisokba a gén kifejeződésének helyeként a petesejt és/vagy a zigóta.

Az cDNS szekvenálásból származó szekvencia információ egyik lehetséges felhasználási lehetősége a „DNS chip” vagy „DNS mikroarray” készítése és átfogó génkifejeződési változások nyomon követése. A búza esetében, mint korábban említésre került, teljes genomot reprezentáló DNS chip még nem készülhetett el, hiszen a búza genom szekvenálása még nem fejeződött be. A legrepresentatívabb búza DNS chip-et az Affymetrix cég (www.affymetrix.com) állítja elő és árulja. Ez a DNS „array” 55052 transzkriptumot reprezentál a termesztett búza mind a három genomját képviselve. A „DNS chip” rövid (25 bp), génspecifikus oligonukleotid szekvenciákat hordoz, amely

biztosítja a magas fokú specificitást, de speciális leolvasó készüléket igényel. Ez számunkra jelenleg nem hozzáférhető. Az Affymetrix búza DNS „chip” hátránya még, hogy drága. Egyetlen „chip” ára 500 Euro körül mozog.

A kutatók, hogy a költségeket csökkentsék és szélesebb körben elérhetővé tegyék a technikát, saját DNS „chip”-ek fejlesztésébe kezdtek. Az Egyesült Királyságban 35 különböző búza cDNS bankból összesen mintegy 10000 cDNS szekvenciát izoláltak és vittek fel DNS „chip”-re (Wilson és mtsai. 2004). Ebben az esetben, a teljes cDNS szekvenciák használata miatt, a specificitás foka alacsonyabb, mint oligonukleotidok alkalmazásakor. Ez a DNS chip kutatási célokra költség áron használható, de a kísérleteket az angliai laboratóriumban lehet csak elvégezni/elvégeztetni.

Japánban, egy búza genomikai konzorcium, 22000 búza gént reprezentáló oligonukleotid „chip”-et hozott létre, 60-70 bp hosszúságú génspecifikus oligonukleotidokat szintetizáltatva (Kawaura és mtsai. 2006). Ez a megoldás a két fentebb említett DNS „chip” kompromisszumának tekinthető: a specificitás mértéke a kettő közé tehető, de a leolvasás olcsóbb készülékekkel is megoldható. A japán kutatók sikerrel használták ezt az oligonukleotid chip-et só stressz hatására aktiválódó/csendesedő búza gének azonosítására (Kawaura és mtsai. 2006).

Laboratóriumunkban praktikus okokból mi is ezt az utóbbi megközelítést használtuk egy búza DNS „chip” előállítására. Célunk az volt, hogy egy viszonylag olcsó, így akár a nemesítők által is használható, speciális búza DNS „mikroarray”-t hozzunk létre, amely alkalmas a búza szemfejlődésével kapcsolatos génkifejeződési változások vizsgálatára. Ehhez mintegy 1500-2000 olyan búza cDNS szekvenciát választottunk ki az adatbázisokból, amelyek kizárólag vagy nagyobb gyakorisággal a búza szemfejlődéshez kapcsolódó cDNS bankokban fordulnak elő. Ezek a cDNS-ek reprezentálják a megtermékenyítéstől a szem feltöltődéséig szerepet játszó géneket, ezen belül a legfontosabb anyagcsere utak és szabályozási mechanizmusok génjeit. Természetesen az elkészült DNS „chip” közel sem tartalmazza a szemfejlődés rendkívül komplex folyamatában szerepet játszó valamennyi gént, de bízunk abban, hogy hazai és külföldi együttműködések révén a DNS „chip” továbbfejleszthető lesz. Ennek első jele, hogy a SZBK „Sejtosztódási ciklus és stressz adaptáció” kutató csoportjában folyó kutatások eredményeként mintegy 1500 oligonukleotid készült el, amelyek a búza környezeti

stresszválaszában szerepet játszó géneket reprezentálnak. Ezek felkerültek az immár közös DNS „chip”-re, amely így közel 3000 búza gént tartalmaz.

6. 4. Szekvencia adatok kezelése és feldolgozása

A modern „High throughput” technikák alkalmazásával az elérhető információ mennyisége olyan mértékben megnőtt, hogy nagy kapacitású, összetett számítástechnikai rendszerek segítségével nélkül lehetetlen elemezni az adatokat. Ennek érzékeltetésére egy példa: Az átlagos ember kb. 2-3000 szót használ a mindennapi beszédben és kb. 2-30000-t ismer. Ez nagyságrendileg megegyezik az ember vagy rizs genomjában található gének számával. Ha ehhez hozzátesszük, hogy az egyedek alkotó több száz szövet, szerv transzkripciós profilja mind különbözik egymástól és ez a profil folyamatosan változik az az egyed fejlődés során vagy különböző külső ingerek hatására (szárazság stressz, gyógyszerkezelés), akkor könnyen beláthatjuk, hogy a feltett kérdések megválaszolásához már nem elég a papír-ceruza módszer.

Mostanában egyre inkább követelmény, hogy a gének funkcióját, ne csak egy specializált kísérletben vizsgáljuk, hanem általános következtetéseket is levonjunk, beleillesztve az újonnan megszerzett tudást az eddigi ismeretekbe. Ehhez elengedhetetlen, hogy a nagyszámú speciális adatbázis (3D szerkezet, transzkripciós profil, kölcsönható gének, fehérje motívum, szignál transzdukciós utak stb.) között a kapcsolatok megléte. Létrehoztam egy olyan általános célú adatbázis-kezelő programot, mely számos a molekuláris biológiában gyakran használt fájl formátumot ismer. A program lehetővé teszi, hogy az adatbázis nyelvek használatában járatlan felhasználó is egyszerűen, grafikus felületen, képes legyen SQL lekérdezéseket tervezni. A program úgyszintén támogatja új adattáblák létrehozását, mind lekérdezések eredményéből, mind új adatok importjából. Speciális funkció, hogy támogatja a KEGG hierarchikus és „Gene Ontology” adatbázisok kapcsolását is. Ezen kívül olyan specifikus funkciók vannak beépítve, melyek megkönnyítik a még nem szekvenált organizmusok EST-éiből készített génlisták elemzését, szekvencia adatokból oligonukleotidok tervezését, elemzését. Lehetőség van új EST-k azonosítására, valamint klónkönyvtárak között a közös EST-k megkeresésére is.

7. Összefoglalás

Munkánk során elkészítettem egy általánosan használható bioinformatikai programot. A szoftver képes számos biológiai adatbázis adatainak, vagy biológia programok eredményeinek beolvasására:

- KEGG
- NCBI GenBank
- TIGR „Gene Indices”
- TIGR „Transcript Assemblies”
- PICKY
- BLAST m7 (XML) formátum
- BLAST m9 (tabulátorral tagolt)
- GO (XML) formátum
- FASTA

Ezen kívül képes bármilyen tabulárisan tagolt fájlból a felhasználó által megadott szabályok alkalmazásával adattáblát készíteni és az adatbázisba elmenteni. Alkalmas arra is, hogy az „SQL” szintaxisban járatlan felhasználó is képes legyen számára fontos, akár összetett lekérdezéseket végezni és az eredményt adatbázisba, vagy EXCEL által olvasható tabulátorral tagolt fájlba elmenteni.

Megterveztem 1504 olyan oligonukleotidot, amely kizárólag olyan géneket ismer fel, amelyek a búzaszem fejlődése folyamán, a szem valamelyik szövetében fejeződnek ki. Az oligonukleotidok felhasználásával készülő microarray tesztelésében és optimalizálásában vettem részt.

15000 oligonukleotidot tartalmazó micro array-t terveztem, amely alkalmas a búzaszemben bekövetkező expressziós változások részletes nyomon követésére.

Létrehoztam egy nagyon jó minőségű búza pete cDNS klónkönyvtárat, és egy ennél valamivel gyengébb, de még mindig jónak mondható 8 órás zigóta könyvtárat.

Megszekvenáltattunk mintegy 700 cDNS klónt és elvégeztem az így kapott EST könyvtárak bioinformatikai elemzését. Meghatároztam számos eddig ismeretlen búza gén szekvencia részletét.

Azonosítottam néhány pete specifikusan, másrészt a megtermékenyítést követően a zigótában erősen expresszáladó gént.

8. Köszönetnyilvánítás

Szeretném megköszönni Dr. Fehér Attilának, témavezetőmnek, mindazt, amit akaratom ellenére elkövetett, hogy ez az értekezés előbb vagy utóbb, de végre létrejöjjön. Köszönöm szakmai segítségét, határtalan türelmét.

Köszönöm minden volt és jelenlegi munkatársamnak, hogy elviselt, mi több még segített is: Dorjgotov Duulgun (Dulgi), Dr. Ötvös Krisztina, Domoki Mónika, Bíró Judit, Dunainé Fodor Csilla, Manuela Jurca, Katona Gyöngyvér, Nagy Rózsa, Szabó Katalin. Dr. Györgyey Jánost (Arthur) a kitűnő bioinformatikai meglátásaiért illeti dicséret. Dr. Endre Gabriellának és Dr. Kiss Ernőnek a microarray analízisekhez nyújtott segítségét köszönöm meg. Dr. Kiss Ernőnek mérhetetlen hálával tartozom, hogy ezt az értekezést nagyon rövid idő alatt hajlandó volt átolvasni és elbírálni a házivédésre.

Hálával tartozom családomnak, barátaimnak, hogy mellettem vannak, és elnézést kérek az ellopott időért azoktól, akik váltig állították, hogy ez az egész butaság, vannak fontosabb dolgok is, és hetek óta nyaggatnak, hogy jobban tenném, ha felmondanék, azaz a gyermekeimtől.

9. Hivatkozási lista

Alexandersson E., Frayse L., Sjövall-Larsen S., Gustavsson S., Fellert M., Karlsson M., Johanson U. & Kjellbom P. (2005) Whole gene family expression and drought stress regulation of aquaporins. *Plant Molecular Biology* 59, 469-484.

Andersen M.N., Ash F., Jensen C.R., Naested H., Mogensen V.O. & Koch K.E. (2002) Soluble invertase expression is an early target of drought stress during the critical, abortion-sensitive phase of young ovary development in maize. *Plant Physiology* 130, 591-604.

Appels, R., Francki, M., Chibbar, R. (2003) Advances in cereal functional genomics. *Funct. Integr. Genomics* 3, 1-24.

Araus J.L., Slafer G.A., Reynolds M.P. & Royo C. (2002) Plant breeding and drought in C3 cereals: what should we breed for? *Annals of Botany* 89, 925-940.

Ash F., Andersen M.N., Jensen C.R. & Mogensen V.O. (2001) Ovary abscisic acid concentration does not induce kernel abortion in field-grown maize subjected to drought. *European Journal of Agronomy* 15, 119-129.

Bahieldin A., Mafouz H.T., Eissa H.F., Saleh O.M., Ramadan A.M., Ahmed I.A., Dyer W.E., El-Itriby H.A. & Makour M.A. (2005) Field evaluation of transgenic wheat plants stably expressing the HVA1 gene for drought tolerance. *Physiologia Plantarum* 123, 421-427.

Barnabás B., Pónya Z., Szakács E., Tímár I., Obert B. & Pretova A. (2001) Biotechnology and micromanipulation of sexual processes in flowering plants. *Biologia* 56, 7-12.

Bartels D. & Souer E. (2004) Molecular responses of higher plants to dehydration. In Plant Responses to Abiotic Stress (eds H. Hirt & K. Shinozaki), pp. 9-38. Springer-Verlag, Berlin & Heidelberg, Germany.

Bennett, M. D., Smith, J. B. (1976) Nuclear-DNA amounts in Angiosperms. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 274, 227-274.

Brandt S., Kloska S., Altmann T. & Kehr J. (2002) Using array hybridization to monitor gene expression at the single cell level. Journal of Experimental Botany 53, 2315-2323.

Bruce W.B., Edmeades G.O. & Barker T.C. (2002) Molecular and physiological approaches to maize improvement for drought tolerance. Journal of Experimental Botany 53, 13-25.

Burke J.J., O'Mahony P.J. & Oliver M.J. (2000) Isolation of Arabidopsis mutants lacking components of acquired thermotolerance. Plant Physiology 123, 575–587.

Busch W., Wunderlich M. & Schöffl F. (2005) Identification of novel heat shock factor-dependent genes and biochemical pathways in *Arabidopsis thaliana*. The Plant Journal 41 (1), 1–14.

Chen W., Provart N.J., Glazebrook J., Katagiri F., Chang H., Eulgem T., Mauch F., Luan S., Zou G., Whitham S.A., Budworth P.R., Tao Y., Xie Z., Chen X., Lam S., Kreps J.A., Harper J.F., Si-Ammour A., Mauch-Mani B., Heinlein M., Kobayashi K., Hohn T., Dangel J.L., Wang X. & Zhu T. (2002) Expression profile matrix of Arabidopsis transcription factor genes suggests their putative functions in response to environmental stresses. Plant Cell 14, 559-574.

Chen W.J. & Zhu T. (2004) Networks of transcription factors with roles in environmental stress response. Trends in Plant Sciences 9, 591-596.

Cominelli E., Galbiati M., Vavasseur A., Conti L., Sala T., Vuylsteke M., Leonhardt N., Dellaporta S.L. & Tonelli C. (2005) A guard-cell-specific MYB transcription factor regulates stomatal movements and plant drought tolerance. *Current Biology* 15,1196-1200.

Craufurd P.Q. & Peacock J.M. (1993) Effect of heat and drought stress on sorghum (*Sorghum bicolor*). II. Grain yield. *Experimental Agriculture* 29, 77-86.

Dupont F. & Altenbach S. (2003) Molecular and biochemical impacts of environmental factors on wheat grain development and protein synthesis. *Journal of Cereal Science* 38, 133-146.

Engelen-Eigles G., Jones R. J. & Phillips R. L. (2001) DNA endoreduplication in maize endosperm cells is reduced by high temperature during the mitotic phase. *Crop Science* 41, 1114-1121

Gebbing T. & Schnyder H. (1999) Pre-anthesis reserve utilization for protein and carbohydrate synthesis in grains of wheat. *Plant Physiology* 121, 871-878.

Feuillet, C., Keller, B. (2002) Comparative genomics in the grass family: molecular characterization of grass genome structure and evolution. *Ann. Bot. (Lond)* 89, 3-10.

Hajheidari M., Eivazi A., Buchanan B.B., Wong J.H., Majidi I. & Salekdeh G.H. (2007) Proteomics uncovers a role for redox in drought tolerance in wheat. *Journal of Proteome Research* 6, 1451-1460.

Heyne E.G. & Brunson A.M. (1940) Genetic studies of heat and drought tolerance in maize. *Journal of American Society of Agronomy* 32, 803-814.

Hong S.W. & Vierling E. (2000) Mutants of *Arabidopsis thaliana* defective in the acquisition of tolerance to high temperature stress. *Proceedings of the National Academy of Sciences USA* 97, 4392–4397.

Huang, X. and Madan, A. (1999) CAP3: A DNA Sequence Assembly Program. *Genome Research*, 9: 868-877.

Jeanneau M., Gerentes D., Foueillassar X., Zivy M., Vidal J., Toppan A. & Perez P. (2002) Improvement of drought tolerance in maize: towards the functional validation of the *Zm-Asr1* gene and increase of water use efficiency by over-expressing C4-PEPC. *Biochimie* 84, 1127-1135.

Kawaura, K., Mochida, K., Yamazaki, Y., Ogiwara, Y. (2006) Transcriptome analysis of salinity stress responses in common wheat using a 22k oligo-DNA microarray. *Funct. Integr. Genomics* 6, 132-142.

Koltunow A.M., Bicknell R.A., Chaudury AM. (1995) Apomixis: Molecular strategies for the generation of genetically identical seeds without fertilisation. *Plant Physiol.* 108:1345-1352

Koltunow, AM (1993) Apomixis: embryo sacs and embryo formed without meiosis

Koonjul P.K., Minhas J.S., Nunes C., Sheoran I.S. & Saini H.S. (2005) Selective transcriptional down-regulation of anther invertases precedes the failure of pollen development in water-stressed wheat. *Journal of Experimental Botany* 56, 179-190.

Langridge P., Paltridge N. & Fincher G. (2006) Functional genomics of abiotic stress tolerance in cereals. *Briefings of Functional Genomics & Proteomics* 4, 343-354.

Majoul T., Bancel E., Triboï E., Ben Hamida J. & Branlard G. (2003) Proteomic analysis of the effect of heat stress on hexaploid wheat grain: characterization of heat-responsive proteins from total endosperm. *Proteomics* 3, 175-183.

Majoul T., Bancel E., Triboï E., Ben Hamida J. & Branlard G. (2004) Proteomic analysis of the effect of heat stress on hexaploid wheat grain: characterization of heat-responsive proteins from non-prolamins fraction. *Proteomics* 4, 505-513.

Mascarenhas J.P. & Crone D.E. (1996) Pollen and the heat shock response. *Sexual Plant Reproduction* 9, 370-374.

McLaughlin J.E. & Boyer J.S. (2004) Sugar-responsive gene expression, invertase activity and senescence in aborting maize ovaries at low water potentials. *Annals of Botany* 94, 675-689.

Moss G.I. & Downey L.A. (1971) Influence of drought stress on female gametophyte development in corn (*Zea mays* L.) and subsequent grain yield. *Crop Science* 11, 368-372.

Nakshima K. & Yamaguchi-Shinozaki K. (2005) Molecular studies on stress-responsive gene expression in *Arabidopsis* and improvement of stress tolerance in crop plants by regulon biotechnology. *Japan Agricultural Research Quarterly* 39, 221–229

Naumova TN, Matzk F (1998) Differences in the initiation of the zygotic and partenogenetic pathway in the Salmon lines of wheat: ultrastructural studies. *Sex Plant Reprod* 11: 121-130

Nicolas M.E., Gleadow R.M. & Dalling M.J. (1985) Effect of post-anthesis drought on cell-division and starch accumulation in developing wheat grains. *Annals of Botany* 55, 433-444.

Okamoto T., Scholten S., Lörz H. & Kranz E. (2005) Identification of genes that are up- or down-regulated in the apical or basal cell of maize two-celled embryos and monitoring their expression during zygote development by a cell manipulation- and PCR-based approach. *Plant and Cell Physiology* 46, 332-338.

Plaut Z., Butow B.J., Blumenthal C.S. & Wrigley C.W. (2004) Transport of dry matter into developing wheat kernels and its contribution to grain yield under post anthesis water deficit and elevated temperature. *Field Crops Research* 86, 185-198.

Rampino P., Pataleo S., Gerardi C., Mita G. & Perrotta C. (2006) Drought response in wheat: physiological and molecular analysis of resistant and sensitive genotypes. *Plant, Cell & Environment* 29, 2143-2152.

Rizhsky L., Liang H. & Mittler R. (2002) The combined effect of drought stress and heat shock on gene expression in tobacco. *Plant Physiology* 130, 1143-1151.

Russell, S. D. (1993) The egg cell: Development and role in fertilization and early embryogenesis. *Plant Cell* 5, 1349-1359.

Saini H.S. & Aspinall D. (1982) Abnormal sporogenesis in wheat (*Triticum aestivum* L.) induced by short periods of high temperature. *Annals of Botany* 49, 835-846.

Saini H.S. & Lalonde S. (1998) Injuries to reproductive development under water stress, and their consequences for crop productivity. *Journal of Crop Production* 1, 223-248.

Saini H.S. (1997) Effect of water stress on male gametophyte development in plants. *Sexual Plant Reproduction* 10, 67-73.

Saini H.S., Sedgley M. & Aspinall D. (1983) Effect of heat stress during floral development on pollen tube growth and ovary anatomy in wheat (*Triticum aestivum* L.). *Australian Journal of Plant Physiology* 10, 137-144.

Saini H.S., Sedgley M. & Aspinall D. (1984) Developmental anatomy in wheat of male sterility induced by heat stress, water deficit or abscisic acid. *Australian Journal of Plant Physiology* 11, 243-253.

Sasaki, T., Sederoff, R. R. (2003) Genome studies and molecular genetics - The rice genome and comparative genomics of higher plants - Editorial overview. *Curr. Op. Plant Biol.* 6, 97-100.

Savin R. & Nicolas M.E (1996) Effects of short periods of drought and high temperature on grain growth and starch accumulation of two malting barley cultivars. *Australian Journal of Plant Physiology* 23, 201–210.

Seki M., Narusaka M., Ishida J., Nanjo T., Fujita M., Oono Y., Kamiya A., Nakajima M., Enju A., Sakurai T., Satou M., Akiyama K., Taji T., Yamaguchi-Shinozaki K., Carninci P., Kawai J., Hayashizaki Y. & Shinozaki K. (2002) Monitoring the expression profiles of 7000 Arabidopsis genes under drought, cold and high-salinity stresses using a full-length cDNA microarray. *Plant Journal* 31, 279-292.

Setter T. L. & Flannigan B. A. (2001) Water deficit inhibits cell division and expression of transcripts involved in cell proliferation and endoreduplication in maize endosperm. *Journal of Experimental Botany* 52, 1401-1408.

Shinozaki K. & Yamaguchi-Shinozaki K. (1999) Molecular responses to drought stress. In *Molecular Responses to Cold, Drought, Heat and Salt Stress in Higher Plants*. (ed R.G. Landes), pp. 11-28. Austin, TX, U.S.A.

Shrawat A.K. & Lörz H. (2006) Agrobacterium-mediated transformation of cereals: a promising approach crossing barriers. *Plant Biotechnology Journal* 4, 575-603.

Sivamani E., Bahieldin A., Wraith J., Al-Niemi T., Dyer W., Ho T. & Qu R. (2000) Improved biomass productivity and water use efficiency under water deficit conditions in transgenic wheat constitutively expressing the barley HVA1 gene. *Plant Science* 155, 1-9.

Sorrells, M. E., La Rota, M., Bermudez-Kandianis, C. E., Greene, R. A., Kantety, R., Munkvold, J. D., Miftahudin, Mahmoud, A., Ma, X. F., Gustafson, P. J., Qi, L. L. L.,

Echalier, B., Gill, B. S., Matthews, D. E., Lazo, G. R., Chao, S. M., Anderson, O. D., Edwards, H., Linkiewicz, A. M., Dubcovsky, J., Akhunov, E. D., Dvorak, J., Zhang, D. S., Nguyen, H. T., Peng, J. H., Lapitan, N. L. V., Gonzalez-Hernandez, J. L., Anderson, J. A., Hossain, K., Kalavacharla, V., Kianian, S. F., Choi, D. W., Close, T. J., Dilbirli, M., Gill, K. S., Steber, C., Walker-Simmons, M. K., McGuire, P. E., Qualset, C. O. (2003) Comparative DNA sequence analysis of wheat and rice genomes. *Genome Res.* 13, 1818-1827.

Sreenivasulu N., Sopory S.K. & Kavi Kishor P.B. (2007) Deciphering the regulatory mechanisms of abiotic stress tolerance in plants by genomic approaches. *Gene* 388, 1-13.

Stone P. (2001) The effects of heat stress on cereal yield and quality. In *Crop Responses and Adaptations to Temperature Stress*. (ed A.S. Basra), pp. 243-291. Food Products Press, Binghamton, New York.

Vasil I.K. (2007) Molecular genetic improvement of cereals: transgenic wheat (*Triticum aestivum* L.). *Plant Cell Reports* (ahead of print) DOI 10.1007/s00299-007-0338-3.

Verdoy D., Lucas M.M., Manrique E., Covarrubias A.A., De Felipe R. & Pueyo J.J. (2004) Differential organ-specific response to salt stress and water deficit in nodulated bean (*Phaseolus vulgaris*). *Plant, Cell & Environment* 27, 757-767.

Wallwork M.A.B., Jenner C.F., Logue S.J. & Sedgley M. (1998) Effect of high temperature during grain-filling on the structure of developing and malted barley grains. *Annals of Botany* 82, 587-599.

Ware, D. H., Jaiswal, P., Ni, J., Yap, I. V., Pan, X., Clark, K. Y., Teytelman, L., Schmidt, S. C., Zhao, W., Chang, K., Cartinhour, S., Stein, L. D., McCouch, S. R. (2002) Gramene, a tool for grass genomics. *Plant Physiol.* 130, 1606-1613.

Wilson, I. D., Barker, G. L. A., Beswick, R. W., Shepherd, S. K., Lu, C. G., Coghill, J. A., Edwards, D., Owen, P., Lyon, R., Parker, J. S., Lenton, J. R., Holdsworth, M. J., Shewry,

P. R., Edwards, K. J. (2004) A transcriptomics resource for wheat functional genomics. *Plant Biotech. J.* 2, 495-506.

Xiao B., Huang Y., Tang N. & Xiong L. (2007) Over-expression of a LEA gene in rice improves drought resistance under the field conditions. *Theoretical and Applied Genetics* 115, 35-46.

Yamakawa H., Hirose T., Kuroda M. & Yamaguchi T. (2007) Comprehensive expression profiling of rice grain filling-related genes under high temperature using DNA microarray. *Plant Physiology* 144, 258-277.

Yang J.C. & Zhang J.H. (2006) Grain filling of cereals under soil drying. *New Phytologist* 169, 223-236.

Yu, H. J., Hogan, P., Sundaresan, V. (2005) Analysis of the female gametophyte transcriptome of Arabidopsis by comparative expression profiling. *Plant Physiol.* 139, 1853-1869.

Young T.E., Ling J., Geisler-Lee C.J., Tanguay R.L., Caldwell C. & Gallie D.R. (2001) Developmental and thermal regulation of the maize heat shock protein, HSP101. *Plant Physiology* 127, 777-791.

Zhang J.Z., Creelman R.A. & Zhu J.K. (2004) From laboratory to field. Using information from Arabidopsis to engineer salt, cold, and drought tolerance in crops. *Plant Physiology* 135, 615–621.

Zhou J., Wang X., Jiao Y., Qin Y., Liu X., He K., Chen C., Ma L., Wang J., Xiong L., Zhang Q., Fan L. & Deng X.W. (2007) Global genome expression analysis of rice in response to drought and high-salinity stresses in shoot, flag leaf, and panicle. *Plant Molecular Biology* 63, 591-608

Zhu X., Gong H., Chen G., Wang S. & Zhang C. (2005) Different solute levels in two spring wheat cultivars induced by progressive field water stress at different developmental stages. *Journal of Arid Environments* 62, 1-14.

10. Ph. D. thesis

Background

Wheat (*Triticum* sp.) is one of the most important crops in the world. It is cultivated from Helsinki (Finland) to Cape Town (South Africa). The quantity and quality of the wheat yield depends on local conditions (daily oscillation of temperature, the length of daytime, the type of soil, the distribution of rainfall, etc.) In Hungary, wheat is cultivated in 1,000,000 ha/year and a typical yield is 4 ton/ha. The value of the product is usually more than 100 billion HUF/year.

From the last century, the average temperature of the world continuously increases. It is not clear yet that human activity or some natural processes are the main reasons of this phenomenon. The “global warming” does not mean that every regions of the world become warmer, but means the aggregate of local changes in weather. The most probable prediction for Hungary is that the summer will be hotter and droughtier. After the problem was recognized, several research projects were established to better understand wheat genetics underlying stress tolerance in order to promote the development of new cultivars with improved yield safety.

My Ph.D. thesis describes our genomic research approaches which can be used for identifying those genes that are responsible for the quantity and quality of wheat harvest and play role in yield safety. We primarily concentrated our efforts on the drought and heat tolerance of the early phases of grain development in wheat. We hope that in a long term our results will help to accelerate wheat breeding and improve the efficiency of the selection of new cultivars.

Transcription profiling of the wheat egg cell

The life cycle of plants can be divided into a diploid sporophyte and a haploid gametophyte phase. Most of our knowledge comes from the dominant sporophyte cycle, because the gametophyte consists of only few cells which are encapsulated deeply in the

tissues of the sporophyte generation. The new modern methods enable us to investigate this hidden life cycle in plants. With T-DNA insertion mutagenesis more than a hundred mutations were identified which affect the developing gametophyte (Pagnussat et al., 2005). Using DNA micro array, 225 genes were identified as gametophyte specific (Yu et al., 2005).

The egg cell has a special role in the gametophyte, because after the fertilisation the new sporophyte generation is developing from it. In most of the species studied (mainly animals), maternally stored mRNAs of the egg were identified as involved in the establishment of embryonic axes, diversification of cell types and morphological changes during early embryogenesis. Compared with animals, little is known about transcripts stored in egg cells of flowering seed plants (angiosperms) and activation of the zygotic genome after fertilization. There is only limited information about the biological activity of the egg cell and its biological processes. Investigation of the egg cell transcriptome can give us a comprehensive view about these processes.

The model organism, *Arabidopsis*, is inadequate for the direct studying of the earliest steps in embryogenesis, because of its small size and the difficulties in the isolation of egg cells. On the contrary, in the case of wheat and maize, there are routine methodologies for egg cell isolation. These techniques were partly (wheat) developed in the laboratory of Beáta Barnabás (Agricultural Research Institute, HAS, Martonvásár, Hungary). On the basis of their experiences and technical possibilities, we have made a wheat egg cell cDNA library and investigated that what kind of genes are expressed in this specific cell type. Meanwhile it cleared up that other research groups had also chosen this approach. Sprunck and her colleagues, in 2005, published the identification of 404 egg cell-expressed genes and their primary characterization. The sequences of these EST are available from international databases, so we could compare them with those 251 cDNA clones, which were identified by ourselves. The number of common sequences were small, so we identified 237 new egg cell specific genes. 73 of these are unknown wheat sequences, while 188 sequences were already described from other cDNA libraries. The analysis of both sequence pools suggests that there are high metabolic activities and intensive changes at the protein level in the wheat egg cells, so this cell type is not so quiescent, as it was postulated earlier (Russell 1993).

The apomictic „Salmon” genotype of wheat is capable of parthenogenetic development. Kumlehn and their colleagues (2001) identified more than 500 ESTs from these egg cells, but their detailed description has not been submitted into any freely available databases. We can conclude that up to now 1000-1500 wheat genes have been identified which are expressed in the egg cell. Only our cDNA library was prepared in phagemids and it is the most representative (app. $1.5 \cdot 10^6$ genes) that may allow a more detailed analysis. The size of the two other published libraries are much smaller as they consist only few thousands of ESTs, which are cloned into plasmids.

The transcription profiles of one- and two-celled zygotes

Sprunck and her colleagues (2005) made also a cDNA library from two-celled wheat embryos. They determined 789 EST sequences and based on these data they compared the expression profile of the egg cell and the two-celled zygote. In the dividing zygote, the transcription and the translation were more intensive and, naturally, cell cycle specific genes appeared. From two-celled maize embryos, Okamoto and his colleagues (2005) identified cDNA sequences that were expressed in the basal and/or the apical cell. They demonstrated that there is a major difference between the expression profile of this two cell type. These experiments did not answer, however, the question: At what time does the transition of the transcriptosome occur? Our cDNA library, which was made from zygotes 8 hours following the fertilization may help to solve this problem. Up to now, we have identified 369 EST sequences from this one-celled zygote cDNA library. Several transcription factors and proteins participating in signal transduction appeared, which may indicate that the zygote's own gene set is already active at this time. The explicit verification of the early activation of a zygote specific gene set needs additional experiments. 81 of the identified 369 zygote ESTs are new wheat genes or alleles. 23 of them also exist in the egg cell, 22 of them were identified in the 2-days-old zygote cDNA library (Sprunck et al. 2005). The amount of the available sequences are not sufficient for a comprehensive analysis, so the identification of new transcripts should be continued in order to get a better picture on fertilization induced changes in transcription.

Genomic approaches in wheat research

As it was mentioned earlier, wheat (*Triticum aestivum* L.) is one of our most important agricultural crops, therefore the investigation of this species has big importance. The application of modern genetic methods, however, are not so easy, because the genome size of wheat is huge, the chromosomes are colossal and the level of polyploidy is high. The allohexaploid wheat genome consist of 16 Gbase of nucleotides (Bennett and Smith, 1976), while the genome of the model organism rice consists of only 450 Mbase (Sasaki and Sederoff 2003). The main difference between the two genomes comes from the repetitive, non-coding, DNA sequences. The gene sets of these organisms are quite similar to each other (Sorells et al. 2003). The rice genome consists of about 40000 genes (Bennetzen et al. 2004). The estimations of the number of genes in wheat are in quite a big range, but surely they are far more than in rice (e.g. considering polyploidy).

The rice and barley as model organisms of cereal research can be useful for wheat research and breeding (Appels et al. 2003; Feuillet and Keller 2002; Ware et al. 2002). Because of the size of the wheat genome and the high number of repetitive sequences, the sequencing of the whole wheat genome will not be completed in the close future. Therefore, the EST sequencing projects have high priority. For this purpose several, cDNA libraries has been constructed representing various developmental stages and responses for different environmental signals. Our laboratory contributes to this effort by the production of egg cell- and zygote-specific cDNA libraries. 154 out of 620 of our EST sequences show no high local similarity to any other already identified wheat EST sequences (less then 90 % percent of identity), so probably they are new, up to now not identified wheat genes or alleles. In several other instances we could extend the already known sequences by our ESTs, while in those cases where the cDNA sequences represented by our ESTs were already fully known, the new information was that these genes are expressed in the egg cell and/or zygote.

One of the potential use of these EST sequences is the production of microarrays and the investigation of comprehensive gene expression patterns. Whole genome microarrays of wheat have not been manufactured yet, due to the unfinished genome sequencing. The most representative microarray for wheat was developed by Affymetrix

(www.affymetrix.com). 55052 different transcripts can be detected by this array. It is using short (25 bp) oligonucleotids which guarantees the high level of specificity. The disadvantages of this method: the requirement of a special scanner and the high price (1 slide app. 500 Euro). In order to decrease the cost of the microarray experiments,, several projects were funded. In the United Kindom, 10000 cDNA sequences from 35 different libraries were amplified using PCR and spotted on microarray slides (Wilson et al. 2004). In this case, the spotted DNA probes were full length cDNA, so this approach is not as specific as the Affymetrix solution. This microarray can be used at a lower price, but the experiments must be done in the UK laboratory. In Japan, a wheat genomic consortium developed an oligonucleotid microarray, which represents 22000 different wheat transcripts (Kawaura et al. 2006) and successfully applied it for identifying genes the expression of which is modified by salt stress.

Our objective was to design and manufacture a relatively cheap, wheat-specific oligonucleotid array (oligo-chip), which can be used even by breeders. We also intended to use this array to investigate the changes of the wheat transcriptosom in the grain, during the grain filling period. We chose about 2000 cDNA sequences from publicly available databases, which exist only or very abundantly in cDNA libraries associated with grain development. Of course, the prepared microarray does not represent the whole gene set which participates in the complex grain filling process. However, it may give a sufficiently detailed picture about the processes underlying this important event. One of our collaborators (BRC, Cell Division Cycle and Stress Adaptation Group) designed additional 1500 oligonucleotides representing stress-responsive genes in wheat. The final oligochip developed in our laborytory therefore contains more than 3000 oligonucleotides.

Handling and processing of large pools of sequence data

The amount of available biological and DNA/protein sequence information is increased so much due to modern high throughput research methods that without the help of complex computer technology the data processing and analysis is impossible. There is also a demand that the new results should be incorporated as a new part into a complex information set. To discover the real relation among several experimental circumstances

and the data obtained by various methodologies, we have to establish connections between several databases (3D modeling, transcription profile, interacting partners, protein motifs etc.)

I have developed a software solution which handles several molecular biology specific file formats (BLAST, PICKY, GenBank, FASTA, etc.) and is capable to handle large amount of data (e.g. 2.2 gigabyte input GenBank file). This program also makes it possible for an inexperienced users to create SQL queries on a graphical interface. The software allows the creation of new data tables by importing tab delimited files or the results of SQL queries. It is also possible to cross-link the data tables to hierarchical databases such as KEGG or Gene Ontology. Several specific functions have been introduced in order to allow the genomic analysis of organisms with not fully sequenced genomes based on their EST sequences and the distribution of their EST sequences in the various cDNA libraries (“virtual gene expression”).

11. Publikációs lista

Dorjgotov D, Szucs A, Otvos K, Szakonyi D, Kelemen Z, Lendvai A, Ponya Z, Barnabas B, Brown S, Dudits D, Feher A.

Specific features of RHO GTPase-dependent signaling in plants.

Cell Biol Int. 2003;27(3),191-2.

impakt faktor: 1.09

Otvos K, Pasternak TP, Miskolczi P, Domoki M, Dorjgotov D, Szucs A, Bottka S, Dudits D, Feher A. Nitric oxide is required for, and promotes auxin-mediated activation of, cell division and embryogenic cell formation but does not influence cell cycle progression in alfalfa cell cultures.

Plant J. 2005 Sep;43(6),849-60.

impakt faktor: 6.37

Szucs A, Dorjgotov D, Otvos K, Fodor C, Domoki M, Gyorgyey J, Kalo P, Kiss GB, Dudits D, Feher A.

Characterization of three Rop GTPase genes of alfalfa (*Medicago sativa* L.).

Biochim Biophys Acta. 2006 Jan-Feb;1759(1-2),108-15.

impakt faktor: 2.04

Szűcs A, Ignáth I, Jäger K, Barnabás B, Fehér A (2007) Fertilization induces gene expression changes in wheat egg cells within 8 hours. Plant Cell Reports (in preparation)

12. Függelék

12. 1. Nagy homológiát mutató pete EST-k annotációja

név	annotáció	e érték
102-tex.	Hypothetical protein P0030H07.38 [Oryza sativa (japonica cultivar-group)]	1.87E-17
103-tex.	14-3-3 protein [Hordeum vulgare (Barley)]	0
105-tex.	contains ESTs AU097236(S3106),D40915(S3106)~similar to endo-beta-glucuronidase/heparanase [Oryza sativa (japonica cultivar-group)]	0
106-tex.		0
110-tex.	Ribosomal protein L15 [Oryza sativa (japonica cultivar-group)]	0
111-tex.	40S ribosomal protein S21 [Oryza sativa (Rice)]	0
112-tex.	Calreticulin-like protein [Triticum aestivum (Wheat)]	0
113-tex.	Zinc finger POZ domain protein-like [Oryza sativa (japonica cultivar-group)]	7.14E-09
115-tex.		2.35E-63
116-tex.	P0402A09.25 protein [Oryza sativa (japonica cultivar-group)]	5.14E-134
118-tex.	Epsilon1-COP [Oryza sativa (japonica cultivar-group)]	0
119-tex.	T6D22.2 [Arabidopsis thaliana (Mouse-ear cress)]	0
120-tex.	Histone deacetylase 2b [Zea mays (Maize)]	0
121-tex.	putative zinc protease [Oryza sativa (japonica cultivar-group)]	3.73E-75
126-tex.	Putative hydroxymethyltransferase [Oryza sativa (japonica cultivar-group)]	6.80E-67
128-tex.	Armadillo/beta-catenin repeat protein-like [Oryza sativa (japonica cultivar-group)]	0
129-tex.	Putative thioredoxin [Oryza sativa (japonica cultivar-group)]	1.09E-62
130-tex.	40S ribosomal protein S8 [Oryza sativa (Rice)]	0
131-tex.	Eukaryotic initiation factor 4A [Triticum aestivum (Wheat)]	0
133-tex.	(BTB-POZ AND MATH DOMAIN 1); protein binding [Arabidopsis thaliana]	1.55E-44
134-tex.	Hexose transporter [Zea mays (Maize)]	0
137-tex.	Nascent polypeptide-associated complex NAC; UBA-like [Medicago truncatula (Barrel medic)]	0
140-tex.	Hypothetical protein OSJNBb0072E24.9 [Oryza sativa (Rice)]	0
141-tex.	40S ribosomal protein S21 [Oryza sativa (Rice)]	0
142-tex.	Mitochondrial chaperonin-60 [Oryza sativa (japonica cultivar-group)]	0
143-tex.	Translation initiation factor eIF-5A [Oryza sativa (japonica cultivar-group)]	0
144-tex.	Putative tumor differentially expressed protein 1 [Oryza sativa (japonica cultivar-group)]	0
149-tex.	similar to 60S ribosomal protein L23a	2.58E-167
150-tex.	Glyceraldehyde-3-phosphate dehydrogenase, cytosolic [Hordeum vulgare (Barley)]	0
151-tex.	Cytosolic heat shock protein 90 [Hordeum vulgare (Barley)]	0
152-tex.		1.28E-81
153-tex.		5.36E-61
158-tex.	Hypothetical protein OSJNBa0003O19.4 [Oryza sativa (Rice)]	0
159-tex.	Putative ribosomal protein L37 [Oryza sativa (japonica cultivar-group)]	2.12E-92
15-tex.n	PREDICTED: similar to 60S ribosomal protein L23a [Homo sapiens]	1.14E-131

161-tex	[Oryza sativa (japonica cultivar-group)]	3.73E-14
163-tex	[Oryza sativa (japonica cultivar-group)]	4.25E-18
166-tex.		4.29E-13
169-tex.	Expressed protein [Oryza sativa (japonica cultivar-group)]	0
16-tex.n	OSJNBa0058G03.7 protein [Oryza sativa (japonica cultivar-group)]	6.45E-127
170-tex	[Oryza sativa (japonica cultivar-group)]	2.10E-41
172-tex.	Similarity to intracellular protein [Arabidopsis thaliana (Mouse-ear cress)]	0
174-tex.	Hydrolase-like protein [Oryza sativa (japonica cultivar-group)]	7.33E-71
175-tex.	OSJNBb0070J16.13 protein [Oryza sativa (japonica cultivar-group)]	0
177-tex.		0
178-tex.	NADH dehydrogenase 10.5K chain-like protein [Arabidopsis thaliana (Mouse-ear cress)]	7.24E-133
179-tex.	Ribosomal RNA apurinic site specific lyase [Triticum aestivum (Wheat)]	5.15E-175
17-tex.n	Hypothetical protein P0519E02.19 [Oryza sativa (japonica cultivar-group)]	0
180-tex.	Zinc finger CCCH type domain-containing protein ZFN-like 2 [Oryza sativa (Rice)]	0
182-tex.	Ubiquitin-like protein SMT3 [Oryza sativa (Rice)]	0
183-tex.	Ribosomal protein L15 [Oryza sativa (japonica cultivar-group)]	0
186-tex.		0
187-tex.	Putative phosphoribosylanthranilate transferase [Oryza sativa (japonica cultivar-group)]	3.00E-126
188-tex.	Probable methionyl-tRNA synthetase [Oryza sativa (Rice)]	1.18E-09
18-tex.n		2.40E-11
190-tex.	Putative OsCTTP [Oryza sativa (japonica cultivar-group)]	2.87E-179
193-tex.	Ribosomal protein L18a-like [Oryza sativa (japonica cultivar-group)]	0
196-tex.	Cytosolic glyceraldehyde-3-phosphate dehydrogenase [Triticum aestivum (Wheat)]	0
197-tex.	ECA1 protein [Hordeum vulgare (Barley)]	3.67E-98
198-tex.	Hypothetical protein B1130G10.20 [Oryza sativa (japonica cultivar-group)]	1.17E-87
199-tex.	Ubiquitin C variant [Homo sapiens (Human)]	7.59E-109
19-tex.n		0
203-tex.	SUI1 protein [Coffea arabica (Coffee)]	0
204-tex.	Galactonolactone dehydrogenase [Oryza sativa (japonica cultivar-group)]	5.44E-118
207-tex.	Hypothetical protein T12C14_220 [Arabidopsis thaliana (Mouse-ear cress)]	6.14E-10
208-tex.	Hypothetical protein OJ1699_E05.34 [Oryza sativa (japonica cultivar-group)]	3.85E-69
209-tex.	Phosphoglycerate kinase, cytosolic [Triticum aestivum (Wheat)]	0
20-tex.n		2.89E-23
210-tex.	GTP-binding nuclear protein Ran/TC4 [Vicia faba (Broad bean)]	0
211-tex.	Chromosome 9 SCAF14729, whole genome shotgun sequence [Tetraodon nigroviridis (Green puffer)]	0
212-tex.	NME2 protein [Homo sapiens (Human)]	1.03E-150
214-tex.	H0423H10.8 protein [Oryza sativa (Rice)]	0
215-tex.	Cytosolic heat shock protein 90 [Hordeum vulgare (Barley)]	0
217-tex.		2.28E-158
219-tex.	Putative alliinase [Oryza sativa (japonica cultivar-group)]	2.99E-129
21-tex.n	60S ribosomal protein L44-like [Arabidopsis thaliana (Mouse-ear cress)]	0
220-tex.	Vacuolar proton-inorganic pyrophosphatase [Hordeum vulgare (Barley)]	0
221-tex.		4.66E-178
222-tex.		7.96E-34
223-tex.	Putative OsCTTP [Oryza sativa (japonica cultivar-group)]	0

224-tex.	Putative ribosomal protein L32 [Oryza sativa (japonica cultivar-group)]	0
226-tex.	40S ribosomal protein S8 [Oryza sativa (Rice)]	0
230-tex.	OSJNBa0084K20.10 protein [Oryza sativa (japonica cultivar-group)]	0
231-tex.		1.67E-104
232-tex.	Putative ribosomal protein [Oryza sativa (japonica cultivar-group)]	0
234-tex.	Putative dihydrolipoamide dehydrogenase [Oryza sativa (japonica cultivar-group)]	4.27E-147
235-tex.		0
236-tex.		0
237-tex.	Hypothetical protein P0030H07.35 [Oryza sativa (japonica cultivar-group)]	7.82E-34
238-tex.	Putative ubiquitin-associated (UBA) protein [Oryza sativa (japonica cultivar-group)]	0
239-tex.	Putative dihydrolipoamide dehydrogenase [Oryza sativa (japonica cultivar-group)]	4.27E-147
240-tex.	putative carbamoyl phosphate synthetase [Oryza sativa (japonica cultivar-group)]	0
242-tex.	PREDICTED P0048D08.116 gene product [Oryza sativa (japonica cultivar-group)]	8.44E-159
244-tex.	putative carbamoyl phosphate synthetase [Oryza sativa (japonica cultivar-group)]	0
245-tex.	Glyceraldehyde-3-phosphate dehydrogenase, cytosolic [Hordeum vulgare (Barley)]	2.29E-142
250-tex.	Ribosomal protein L11 [Triticum aestivum (Wheat)]	2.22E-120
253-tex.	Armadillo/beta-catenin repeat protein-like [Oryza sativa (japonica cultivar-group)]	0
256-tex.	SelfT/selfW/selfH selenoprotein domain, putative [Oryza sativa (japonica cultivar-group)]	0
260-tex.	Hypothetical protein At2g38460 [Arabidopsis thaliana (Mouse-ear cress)]	2.20E-105
262-tex.	Asparagine synthetase [Striga hermonthica (Purple witchweed)]	3.73E-66
263-tex.	Putative ribonucleoprotein [Oryza sativa (japonica cultivar-group)]	0
264-tex.	Putative zinc finger (C3HC4-type RING finger) protein [Oryza sativa (japonica cultivar-group)]	9.65E-167
265-tex.	GTP-binding nuclear protein Ran/TC4 [Vicia faba (Broad bean)]	0
267-tex.	Putative ribosomal protein S29 [Oryza sativa (japonica cultivar-group)]	9.32E-41
269-tex.	Hypothetical protein OSJNBa0089K24.5 [Oryza sativa (japonica cultivar-group)]	3.48E-50
272-tex.	Putative elongation factor 1 beta [Hordeum vulgare (Barley)]	0
275-tex.	Putative ubiquinone oxidoreductase subunit [Oryza sativa (Rice)]	0
27-tex.n	Putative ubiquinone oxidoreductase subunit [Oryza sativa (Rice)]	0
280-tex.	Putative aminopeptidase C [Oryza sativa (japonica cultivar-group)]	5.21E-06
281-tex.		0
29-tex.n	Putative ubiquinone oxidoreductase subunit [Oryza sativa (Rice)]	0
301-tex.	Similar to H. sapiens NNP-1 / Nop52 AP001752 [Hordeum vulgare var. distichum (Two-rowed barley)]	0
309-tex.	Ribosomal protein L36 [Triticum aestivum (Wheat)]	1.09E-178
311-tex.	Putative WD-40 repeat protein [Oryza sativa (japonica cultivar-group)]	0
314-tex.	Hypothetical protein OJ1163_C07.26-2 [Oryza sativa (japonica cultivar-group)]	0
31-tex.n		4.58E-97
320-tex.	OSJNBa0091D06.25 protein [Oryza sativa (Rice)]	5.04E-28
322-tex.	OSJNBa0091D06.25 protein [Oryza sativa (Rice)]	5.04E-28

326-text.	Putative KOW domain-containing transcription factor [Oryza sativa (japonica cultivar-group)]	0
327-text.	Putative 26S proteasome regulatory particle triple-A ATPase subunit3 [Oryza sativa (japonica cultivar-group)]	3.24E-59
328-text.	IDI1 protein [Hordeum vulgare (Barley)]	1.09E-175
330-text.	60S ribosomal protein L19-3 [Arabidopsis thaliana (Mouse-ear cress)]	6.83E-102
332-text.	Putative WD repeat protein [Oryza sativa (japonica cultivar-group)]	0
334-text.	60S acidic ribosomal protein P0 [Zea mays (Maize)]	4.70E-59
335-text.	Putative PHD-finger protein [Oryza sativa (japonica cultivar-group)]	5.34E-65
336-text.	Putative ribosomal protein L34 [Oryza sativa (japonica cultivar-group)]	5.81E-140
337-text.	J2P [Daucus carota (Carrot)]	3.42E-109
339-text.	Nucleosome/chromatin assembly factor group A [Zea mays (Maize)]	9.61E-143
33-text.n	OSJNBa0040D17.11 protein [Oryza sativa (japonica cultivar-group)]	4.12E-21
341-text.	NME2 protein [Homo sapiens (Human)]	3.34E-116
343-text.		3.01E-41
347-text.	Probable splicing factor 3 subunit 1 [Arabidopsis thaliana (Mouse-ear cress)]	0
348-text.		1.13E-133
349-text.		1.53E-100
34-text.n	40S ribosomal protein S11 [Euphorbia esula (Leafy spurge)]	3.95E-65
350-text.	Skp1 [Medicago sativa (Alfalfa)]	4.64E-153
351-text.	Single-stranded nucleic acid binding protein [Triticum aestivum (Wheat)]	4.45E-164
352-text.	Protein disulfide isomerase [Zea mays (Maize)]	0
357-text.		5.54E-34
359-text.	Ribosomal protein L6 [Triticum aestivum (Wheat)]	0
362-text.	Glyceraldehyde-3-phosphate dehydrogenase, cytosolic 3 [Zea mays (Maize)]	0
364-text.	Putative S-adenosyl-L-methionine synthetase [Oryza sativa (japonica cultivar-group)]	0
365-text.		3.33E-94
367-text.		0
368-text.	putative protein kinase [Oryza sativa (japonica cultivar-group)]	2.35E-63
36-text.n	T6D22.2 [Arabidopsis thaliana (Mouse-ear cress)]	1.29E-144
371-text.	40S ribosomal protein S8 [Oryza sativa (Rice)]	0
375-text.	Vacuolar proton-inorganic pyrophosphatase [Triticum aestivum (Wheat)]	0
376-text.	Putative 40S ribosomal protein S3 [Triticum aestivum (Wheat)]	0
377-text.	Nucleotide pyrophosphatase/phosphodiesterase [Hordeum vulgare (Barley)]	0
379-text.		
383-text.	Putative RNA recognition motif (RRM)-containing protein [Oryza sativa (japonica cultivar-group)]	0
384-text.	Ribosomal protein L18a-like [Oryza sativa (japonica cultivar-group)]	0
385-text.	Hypothetical protein OJ1513_F02.133 [Oryza sativa (japonica cultivar-group)]	2.77E-180
386-text.	Calmodulin [Oryza sativa (japonica cultivar-group)]	0
38-text.n		0
391-text.		0
397-text.		9.34E-07
400-text.	Hypothetical protein T1217.4 [Arabidopsis thaliana (Mouse-ear cress)]	8.04E-155
406-text.	Hypothetical protein P0431G05.7 [Oryza sativa (japonica cultivar-group)]	0
409-text.	OSJNBa0044K18.2 protein [Oryza sativa (japonica cultivar-group)]	0
414-text.	Putative transcription factor [Oryza sativa (japonica cultivar-group)]	0
416-text.		4.03E-150

419-tex.	Alcohol dehydrogenase class III-like protein [Solanum tuberosum (Potato)]	0
420-tex.		1.02E-125
428-tex.	Putative transcription initiation factor [Oryza sativa (japonica cultivar-group)]	0
43-tex.n	Putative glycine-rich protein [Oryza sativa (japonica cultivar-group)]	0
44-tex.n	Expressed protein [Oryza sativa (japonica cultivar-group)]	0
468-tex.		0
469-tex.	Guanine nucleotide-exchange protein GEP2 [Oryza sativa (Rice)]	7.20E-105
470-tex.	60S ribosomal protein L21 [Oryza sativa (japonica cultivar-group)]	0
471-tex.	Hypothetical protein P0017B12.18-1 [Oryza sativa (japonica cultivar-group)]	7.08E-54
472-tex.	Ribosomal protein L18a-like [Oryza sativa (japonica cultivar-group)]	0
476-tex.	Putative dihydrolipoamide dehydrogenase [Oryza sativa (japonica cultivar-group)]	4.27E-147
477-tex	dehydrogenase 1, mitochondrial / lipoamide dehydrogenase 1 (MTLPD1) [Arabidopsis thaliana] >gi 30694221 ref NP_849782.1 dihydrolipoamide dehydrogenase 1, mitochondrial / lipoamide dehydrogenase 1 (MTLPD1) [Arabidopsis thaliana]	1.12E-10
477-tex.	Putative dihydrolipoamide dehydrogenase [Oryza sativa (japonica cultivar-group)]	4.27E-147
481-tex.	Putative 60S ribosomal protein L1 [Oryza sativa (japonica cultivar-group)]	9.56E-170
482-tex	ribosomal protein L4/L1 (RPL4D) [Arabidopsis thaliana]	5.95E-12
482-tex.	Fructokinase-2 [Oryza sativa (Rice)]	9.37E-76
485-tex.	Ubiquitin C variant [Homo sapiens (Human)]	3.83E-99
488-tex.	Cysteine synthase (EC 2.5.1.47) (O-acetylserine sulfhydrylase) (O-acetylserine (Thiol)-lyase) [Oryza sativa (Rice)]	0
490-tex	family protein [Arabidopsis thaliana]	1.85E-21
493-tex	dehydrogenase 1, mitochondrial / lipoamide dehydrogenase 1 (MTLPD1) [Arabidopsis thaliana] >gi 30694221 ref NP_849782.1 dihydrolipoamide dehydrogenase 1, mitochondrial / lipoamide dehydrogenase 1 (MTLPD1) [Arabidopsis thaliana]	4.26E-10
498-tex.	Putative X1 [Oryza sativa (japonica cultivar-group)]	1.30E-51
517-tex	[Oryza sativa (japonica cultivar-group)] >gi 115451091 ref NP_001049146.1 Os03g0177500 [Oryza sativa (japonica cultivar-group)] >gi 115451093 ref NP_001049147.1 Os03g0177900 [Oryza sativa (japonica cultivar-group)] >gi 115451095 ref NP_00104914	1.53E-100
519-tex.	T6D22.2 [Arabidopsis thaliana (Mouse-ear cress)]	0
522-tex	binding / zinc ion binding [Arabidopsis thaliana]	6.91E-69
523-tex.	Ubiquitin C variant [Homo sapiens (Human)]	1.05E-49
525-tex	[Oryza sativa (japonica cultivar-group)]	6.36E-14
526-tex.	70 kDa peptidyl-prolyl isomerase [Triticum aestivum (Wheat)]	0
529-tex.	OSJNBa0088A01.4 protein [Oryza sativa (japonica cultivar-group)]	5.42E-33
530-tex.		2.02E-29
533-tex	(heat shock protein 70B); ATP binding [Arabidopsis thaliana]	1.74E-20
537-tex.	Expressed protein [Oryza sativa (japonica cultivar-group)]	9.80E-169
539-tex	family protein (FBL17) [Arabidopsis thaliana]	7.24E-26
539-tex.		3.79E-46
542-tex.	Putative ubiquitin-associated (UBA) protein [Oryza sativa (japonica cultivar-group)]	0
545-tex	(UBA)/TS-N domain-containing protein / octicosapeptide/Phox/Bemp1 (PB1) domain-containing protein [Arabidopsis thaliana]	7.03E-13
546-tex	(heat shock protein 70); ATP binding [Arabidopsis thaliana]	1.12E-57

549-tex.	Putative hsp70 [Oryza sativa (japonica cultivar-group)]	0
550-tex.	Putative mitochondrial carrier protein [Arabidopsis thaliana (Mouse-ear cress)]	0
556-tex.	Thioredoxin peroxidase [Secale cereale (Rye)]	1.12E-110
55-tex.n	Putative RAD23 protein [Oryza sativa (japonica cultivar-group)]	0
56-tex.n	Hypothetical protein OSJNBa0053E01.13 [Oryza sativa (japonica cultivar-group)]	6.81E-169
58-tex.n	putative formin binding protein [Oryza sativa (japonica cultivar-group)]	0
60-tex.n	Hypothetical protein OJ1513_F02.133 [Oryza sativa (japonica cultivar-group)]	1.71E-172
61-tex.n	T6D22.2 [Arabidopsis thaliana (Mouse-ear cress)]	1.62E-105
62-tex.n	Putative 2-oxoglutarate dehydrogenase, E1 subunit [Oryza sativa (japonica cultivar-group)]	0
63-tex.n	Hypothetical protein OSJNBa0078A17.10 [Oryza sativa (japonica cultivar-group)]	0
65-tex.n	(EMBRYO DEFECTIVE 1989); DNA binding [Arabidopsis thaliana]	1.30E-75
66-tex.n	Ribosomal protein S4 [Oryza sativa (japonica cultivar-group)]	0
69-tex.n	Pyruvate kinase [Oryza sativa (japonica cultivar-group)]	1.30E-137
73-tex.n	T6D22.2 [Arabidopsis thaliana (Mouse-ear cress)]	0
86-tex.n	Enolase 2 [Zea mays (Maize)]	0
87-tex.n	Putative phosphatidic acid phosphatase beta [Oryza sativa (japonica cultivar-group)]	0
94-tex.n	Ubiquitin C variant [Homo sapiens (Human)]	0
98-tex.n	Glyceraldehyde-3-phosphate dehydrogenase, cytosolic 3 [Zea mays (Maize)]	0

12. 2. Nagy homológiát mutató zigóta EST-k annotációja

név	annotáció	e érték
zig108-tex5.ab1 1 522	Putative 60S ribosomal protein L12 [Oryza sativa (japonica cultivar-group)]	0
060711-17_B02_ZIG496-TEX5.ab1 850 0	Cytosolic glyceraldehyde-3-phosphate dehydrogenase [Triticum aestivum (Wheat)]	1.25E-178
060711-17_B03_ZIG401-TEX5.ab1 792 0	Eukaryotic translation initiation factor 5A1 [Triticum aestivum (Wheat)]	0
060711-17_B04_ZIG497-TEX5.ab1 856 0	Apoptosis-related protein [Oryza sativa (indica cultivar-group)]	0
060711-17_B05_ZIG402-TEX5.ab1 850 0	Lactoylglutathione lyase [Oryza sativa (Rice)]	1.79E-87
060711-17_B07_ZIG403-TEX5.ab1 847 0	Putative mitochondrial NAD-dependent malate dehydrogenase [Solanum tuberosum (Potato)]	0

060711-17_B09_ZIG404-TEX5.ab1 858 0 858 ABI	S28 ribosomal protein [Triticum aestivum (Wheat)]	6.36E-118
060711-17_B10_ZIG500-TEX5.ab1 858 0 858 ABI		2.07E-124
060711-17_B11_ZIG405-TEX5.ab1 852 0 852 ABI		1.15E-130
060711-17_B12_ZIG501-TEX5.ab1 851 0 851 ABI	PPR-repeat protein-like [Oryza sativa (japonica cultivar-group)]	8.25E-32
060711-17_B13_ZIG406-TEX5.ab1 857 0 857 ABI	Putative elongation factor 1 beta [Hordeum vulgare (Barley)]	0
060711-17_B14_ZIG502-TEX5.ab1 861 0 861 ABI	Putative c-myc binding protein [Oryza sativa (Rice)]	0
060711-17_B15_ZIG407-TEX5.ab1 850 0 850 ABI	Hypothetical protein OJ2056_H01.18 [Oryza sativa (japonica cultivar-group)]	0
060711-17_B16_ZIG503-TEX5.ab1 857 0 857 ABI	OSJNBb0011N17.20 protein [Oryza sativa (japonica cultivar-group)]	0
060711-17_B17_ZIG408-TEX5.ab1 838 0 838 ABI	T6D22.2 [Arabidopsis thaliana (Mouse-ear cress)]	4.57E-175
060711-17_B19_ZIG409-TEX5.ab1 868 0 868 ABI	Hypothetical protein OJ1513_F02.133 [Oryza sativa (japonica cultivar-group)]	8.67E-122
060711-17_B20_ZIG505-TEX5.ab1 852 0 852 ABI	Hypothetical protein OJ1513_F02.133 [Oryza sativa (japonica cultivar-group)]	6.78E-113
060711-17_B22_ZIG506-TEX5.ab1 855 0 855 ABI	ATP binding / ATP-dependent helicase/ DNA binding / helicase/ nucleic acid binding [Arabidopsis thaliana]	0
060711-17_B23_ZIG411-TEX5.ab1 841 0 841 ABI	T6D22.2 [Arabidopsis thaliana (Mouse-ear cress)]	4.57E-175
060711-17_B24_ZIG507-TEX5.ab1 854 0 854 ABI	Hypothetical protein OJ1513_F02.133 [Oryza sativa (japonica cultivar-group)]	8.30E-117
060711-17_D01_ZIG412-TEX5.ab1 666 0 666 ABI	Hypothetical protein OJ1513_F02.133 [Oryza sativa (japonica cultivar-group)]	3.10E-118
060711-	SelT/selW/selH selenoprotein domain, putative [Oryza sativa]	0

17_D04_ZIG509- TEX5.ab1 850 0 850 ABI	(japonica cultivar-group)]	
060711- 17_D05_ZIG414- TEX5.ab1 851 0 851 ABI	ALY protein [Nicotiana benthamiana]	1.89E-21
060711- 17_D06_ZIG510- TEX5.ab1 857 0 857 ABI	Putative Ras-GTPase activating protein SH3 domain-binding protein 2 [Oryza sativa (japonica cultivar-group)]	0
060711- 17_D08_ZIG511- TEX5.ab1 866 0 866 ABI	Lactoylglutathione lyase [Oryza sativa (Rice)]	5.66E-162
060711-17_D09_ZIG416-TEX5.ab1 867 0 867 ABI		1.12E-147
060711- 17_D10_ZIG512- TEX5.ab1 857 0 857 ABI	Ubiquitin-activating enzyme E1-like [Oryza sativa (japonica cultivar-group)]	0
060711- 17_D11_ZIG417- TEX5.ab1 859 0 859 ABI	Hypothetical protein OJ1513_F02.133 [Oryza sativa (japonica cultivar-group)]	2.98E-118
060711- 17_D12_ZIG513- TEX5.ab1 853 0 853 ABI	Small nuclear ribonucleoprotein f, putative [Oryza sativa (japonica cultivar-group)]	0
060711- 17_D14_ZIG514- TEX5.ab1 861 0 861 ABI	Putative FAS1 [Oryza sativa (japonica cultivar-group)]	0
060711- 17_D15_ZIG419- TEX5.ab1 859 0 859 ABI	Putative plastid ribosomal protein S6 [Oryza sativa (japonica cultivar-group)]	0
060711- 17_D16_ZIG515- TEX5.ab1 854 0 854 ABI	Putative X1 [Oryza sativa (japonica cultivar-group)]	5.28E-42
060711-17_D18_ZIG516-TEX5.ab1 856 0 856 ABI		1.13E-166
060711-17_D20_ZIG517-TEX5.ab1 851 0 851 ABI		1.13E-166
060711- 17_D21_ZIG422- TEX5.ab1 865 0 865 ABI	Hypothetical protein P0413G02.21 [Oryza sativa (japonica cultivar-group)]	0
060711- 17_D22_ZIG518- TEX5.ab1 846 0 846 ABI	Hypothetical protein OJ1513_F02.133 [Oryza sativa (japonica cultivar-group)]	2.88E-118
060711- 17_D23_ZIG423- TEX5.ab1 856 0 856 ABI	Progesterone 5-beta-reductase, putative [Oryza sativa (japonica cultivar-group)]	2.92E-111
060711-	Eukaryotic translation initiation factor 4E-1 [Triticum aestivum]	0

17_D24_ZIG519- TEX5.ab1 850 0 850 ABI	(Wheat)]	
060711- 17_F04_ZIG521- TEX5.ab1 853 0 853 ABI	CAA30377.1 protein [Oryza sativa (Rice)]	0
060711- 17_F07_ZIG427- TEX5.ab1 796 0 796 ABI	Glutamine synthetase cytosolic isozyme 1-2 [Oryza sativa (Rice)]	8.97E-46
060711- 17_F08_ZIG523- TEX5.ab1 858 0 858 ABI	Putative ribonucleotide reductase [Oryza sativa (japonica cultivar-group)]	0
060711- 17_F10_ZIG524- TEX5.ab1 853 0 853 ABI	Guanine nucleotide-exchange protein GEP2 [Oryza sativa (Rice)]	1.74E-116
060711- 17_F12_ZIG525- TEX5.ab1 772 0 772 ABI	17.4 kDa class I heat shock protein 3 [Oryza sativa (Rice)]	0.000277
060711- 17_F13_ZIG430- TEX5.ab1 853 0 853 ABI	Minichromosome maintenance protein [Zea mays (Maize)]	3.90E-137
060711- 17_F14_ZIG526- TEX5.ab1 851 0 851 ABI	Ribosomal protein S27-like protein [Solanum tuberosum (Potato)]	2.95E-08
060711-17_F15_ZIG431-TEX5.ab1 853 0 853 ABI		3.42E-110
060711- 17_F18_ZIG528- TEX5.ab1 858 0 858 ABI	Hypothetical protein P0019D06.31 [Oryza sativa (japonica cultivar-group)]	0
060711- 17_F20_ZIG529- TEX5.ab1 856 0 856 ABI	Nucleoside diphosphate kinase 4, chloroplast precursor [Spinacia oleracea (Spinach)]	0
060711-17_F21_ZIG434-TEX5.ab1 854 0 854 ABI		2.38E-78
060711- 17_F23_ZIG435- TEX5.ab1 703 0 703 ABI	Histone H3 [Triticum aestivum (Wheat)]	8.57E-99
060711- 17_F24_ZIG531- TEX5.ab1 839 0 839 ABI	Skp1 [Medicago sativa (Alfalfa)]	3.47E-179
060711- 17_H01_ZIG436- TEX5.ab1 859 0 859 ABI	Histone H3 [Triticum aestivum (Wheat)]	3.84E-29
060711-17_H05_ZIG438-TEX5.ab1 854 0 854 ABI		0
060711-	Skp1 [Medicago sativa (Alfalfa)]	8.36E-66

17_H06_ZIG534- TEX5.ab1 799 0 799 ABI		
060711- 17_H07_ZIG439- TEX5.ab1 852 0 852 ABI	SET domain protein [Triticum aestivum (Wheat)]	8.97E-106
060711- 17_H08_ZIG535- TEX5.ab1 856 0 856 ABI	Putative RNA helicase [Oryza sativa (japonica cultivar-group)]	8.58E-52
060711- 17_H09_ZIG440- TEX5.ab1 850 0 850 ABI	Chromomethylase 1 [Hordeum vulgare (Barley)]	1.35E-131
060711-17_H10_ZIG536-TEX5.ab1 867 0 867 ABI		5.23E-73
060711- 17_H11_ZIG441- TEX5.ab1 865 0 865 ABI	Malate dehydrogenase, cytoplasmic [Zea mays (Maize)]	0
060711- 17_H12_ZIG537- TEX5.ab1 865 0 865 ABI	Hypothetical protein F8F6_270 [Arabidopsis thaliana (Mouse-ear cress)]	0
060711- 17_H13_ZIG442- TEX5.ab1 853 0 853 ABI	Hypothetical protein P0413G02.21 [Oryza sativa (japonica cultivar-group)]	0
060711-17_H14_ZIG538-TEX5.ab1 866 0 866 ABI		1.85E-179
060711- 17_H15_ZIG443- TEX5.ab1 857 0 857 ABI	Ribosomal L10, putative [Oryza sativa (japonica cultivar-group)]	0
060711-17_H16_ZIG539-TEX5.ab1 855 0 855 ABI		4.58E-172
060711- 17_H17_ZIG444- TEX5.ab1 826 0 826 ABI	Histone H2B.2 [Triticum aestivum (Wheat)]	2.00E-130
060711- 17_H18_ZIG540- TEX5.ab1 856 0 856 ABI	Ribosomal protein L30 [Triticum aestivum (Wheat)]	0
060711- 17_H19_ZIG445- TEX5.ab1 870 0 870 ABI	Putative replication protein A2 [Oryza sativa (japonica cultivar-group)]	7.57E-137
060711- 17_H20_ZIG541- TEX5.ab1 859 0 859 ABI	SelT/selW/selH selenoprotein domain, putative [Oryza sativa (japonica cultivar-group)]	0
060711- 17_H21_ZIG446- TEX5.ab1 863 0 863 ABI	Lactoylglutathione lyase [Oryza sativa (Rice)]	3.01E-126
060711-	Lactoylglutathione lyase [Oryza sativa (Rice)]	3.24E-132

17_H22_ZIG542- TEX5.ab1 867 0 867 ABI		
060711- 17_H24_ZIG543- TEX5.ab1 840 0 840 ABI	Prefoldin subunit 3, putative [Oryza sativa (japonica cultivar-group)]	0
060711-17_J01_ZIG448- TEX5.ab1 857 0 857 ABI		6.73E-162
060711- 17_J03_ZIG449- TEX5.ab1 865 0 865 ABI	CUL1 [Oryza sativa (japonica cultivar-group)]	6.60E-12
060711- 17_J04_ZIG545- TEX5.ab1 861 0 861 ABI	UDP-D-glucose epimerase 3 [Hordeum vulgare (Barley)]	0
060711- 17_J05_ZIG450- TEX5.ab1 854 0 854 ABI	S-adenosylmethionine decarboxylase [Narcissus pseudonarcissus (Daffodil)]	5.08E-41
060711- 17_J06_ZIG546- TEX5.ab1 854 0 854 ABI	Hypothetical protein OJ1513_F02.133 [Oryza sativa (japonica cultivar-group)]	1.13E-117
060711- 17_J07_ZIG451- TEX5.ab1 854 0 854 ABI	SeT/seW/seH selenoprotein domain, putative [Oryza sativa (japonica cultivar-group)]	0
060711- 17_J08_ZIG547- TEX5.ab1 844 0 844 ABI	Putative elongation factor 1 beta [Hordeum vulgare (Barley)]	0
060711- 17_J09_ZIG452- TEX5.ab1 851 0 851 ABI	Hypothetical protein OJ1513_F02.133 [Oryza sativa (japonica cultivar-group)]	1.21E-117
060711-17_J10_ZIG548- TEX5.ab1 869 0 869 ABI		0
060711-17_J11_ZIG453- TEX5.ab1 871 0 871 ABI		7.54E-168
060711- 17_J12_ZIG549- TEX5.ab1 834 0 834 ABI	OSJNBa0040D17.15 protein [Oryza sativa (japonica cultivar-group)]	0
060711- 17_J13_ZIG454- TEX5.ab1 861 0 861 ABI	Ubiquitin C variant [Homo sapiens (Human)]	2.76E-175
060711-17_J14_ZIG550- TEX5.ab1 838 0 838 ABI		0
060711- 17_J15_ZIG455- TEX5.ab1 840 0 840 ABI	Spliceosomal protein [Solanum tuberosum (Potato)]	0
060711- 17_J16_ZIG551- TEX5.ab1 818 0 818 ABI	Hypothetical protein P0445D12.10 [Oryza sativa (japonica cultivar-group)]	4.06E-47

060711-17_J17_ZIG456-TEX5.ab1	858	0	858 ABI	3.42E-122
060711-17_J18_ZIG552-TEX5.ab1	840	0	840 ABI	
			Cytosolic heat shock protein 90 [Hordeum vulgare (Barley)]	0
060711-17_J19_ZIG457-TEX5.ab1	848	0	848 ABI	
			60S ribosomal protein L44-like [Arabidopsis thaliana (Mouse-ear cress)]	0
060711-17_J20_ZIG553-TEX5.ab1	848	0	848 ABI	
			Putative polynucleotide adenylyltransferase [Oryza sativa (indica cultivar-group)]	0
060711-17_J21_ZIG458-TEX5.ab1	856	0	856 ABI	
			Spermidine synthase 1 [Oryza sativa (Rice)]	0
060711-17_J22_ZIG554-TEX5.ab1	871	0	871 ABI	
			Disease-resistant-related protein [Oryza sativa (Rice)]	5.75E-94
060711-17_J23_ZIG459-TEX5.ab1	844	0	844 ABI	
			Ribose-phosphate pyrophosphokinase 2 [Oryza sativa (Rice)]	2.45E-62
060711-17_J24_ZIG555-TEX5.ab1	805	0	805 ABI	
			Prefoldin subunit 3, putative [Oryza sativa (japonica cultivar-group)]	1.36E-59
060711-17_L01_ZIG460-TEX5.ab1	860	0	860 ABI	
			Putative plastidic aldolase [Oryza sativa (japonica cultivar-group)]	1.32E-106
060711-17_L02_ZIG556-TEX5.ab1	829	0	829 ABI	
			60s ribosomal protein L21 [Triticum aestivum (Wheat)]	7.35E-96
060711-17_L03_ZIG461-TEX5.ab1	867	0	867 ABI	3.28E-153
060711-17_L04_ZIG557-TEX5.ab1	795	0	795 ABI	
			60S ribosomal protein L44-like [Arabidopsis thaliana (Mouse-ear cress)]	0
060711-17_L05_ZIG462-TEX5.ab1	867	0	867 ABI	
			Ubiquitin C variant [Homo sapiens (Human)]	3.50E-11
060711-17_L06_ZIG558-TEX5.ab1	853	0	853 ABI	
			Lactoylglutathione lyase [Oryza sativa (Rice)]	2.27E-155
060711-17_L07_ZIG463-TEX5.ab1	864	0	864 ABI	
			P0007F06.24 [Oryza sativa (japonica cultivar-group)]	0
060711-			Putative copper chaperone [Oryza sativa (japonica cultivar-	3.57E-167

17_L10_ZIG560- TEX5.ab1 853 0 853 ABI	group))]	
060711- 17_L11_ZIG465- TEX5.ab1 855 0 855 ABI	Putative RNA polymerase sigma factor [Oryza sativa (japonica cultivar-group)]	4.41E-103
060711- 17_L12_ZIG561- TEX5.ab1 867 0 867 ABI	Transport protein subunit-like [Arabidopsis thaliana (Mouse-ear cress)]	0
060711- 17_L13_ZIG466- TEX5.ab1 858 0 858 ABI	Histone H3 [Triticum aestivum (Wheat)]	1.50E-47
060711- 17_L14_ZIG562- TEX5.ab1 855 0 855 ABI	SelT/selW/selH selenoprotein domain, putative [Oryza sativa (japonica cultivar-group)]	0
060711- 17_L15_ZIG467- TEX5.ab1 871 0 871 ABI	26S proteasome AAA-ATPase subunit RPT5a [Arabidopsis thaliana (Mouse-ear cress)]	1.57E-118
060711- 17_L16_ZIG563- TEX5.ab1 857 0 857 ABI	Disease-resistant-related protein [Oryza sativa (Rice)]	6.40E-96
060711-17_L17_ZIG468- TEX5.ab1 858 0 858 ABI		4.49E-61
060711- 17_L18_ZIG564- TEX5.ab1 854 0 854 ABI	Ribosomal protein l34 [Triticum aestivum (Wheat)]	0
060711- 17_L19_ZIG469- TEX5.ab1 794 0 794 ABI	Protein mago nashi homolog [Oryza sativa (Rice)]	9.58E-46
060711- 17_L20_ZIG565- TEX5.ab1 856 0 856 ABI	Hexose transporter [Zea mays (Maize)]	0
060711- 17_L21_ZIG470- TEX5.ab1 911 0 911 ABI	Syntaxin related protein [Oryza sativa (indica cultivar-group)]	0
060711- 17_L22_ZIG566- TEX5.ab1 857 0 857 ABI	Hypothetical protein OJ1513_F02.133 [Oryza sativa (japonica cultivar-group)]	4.36E-117
060711-17_L23_ZIG471- TEX5.ab1 848 0 848 ABI		1.28E-72
060711- 17_L24_ZIG567- TEX5.ab1 853 0 853 ABI	Putative GATA zinc finger protein [Oryza sativa (japonica cultivar-group)]	3.37E-45
060711- 17_N01_ZIG472-	Hypothetical protein OSJNBa0065C16.4 [Oryza sativa (Rice)]	6.97E-159

TEX5.ab1 857 0 857 ABI		
060711-17_N02_ZIG568-TEX5.ab1 855 0 855 ABI	Ribosomal protein S4 [Oryza sativa (japonica cultivar-group)]	5.32E-155
060711-17_N03_ZIG473-TEX5.ab1 856 0 856 ABI	Hypothetical protein OSJNBa0065C16.4 [Oryza sativa (Rice)]	6.97E-159
060711-17_N05_ZIG474-TEX5.ab1 875 0 875 ABI	Putative rab1 small GTP-binding protein [Oryza sativa (japonica cultivar-group)]	0
060711-17_N07_ZIG475-TEX5.ab1 850 0 850 ABI	Putative CCAAT-binding transcription factor [Oryza sativa (japonica cultivar-group)]	0
060711-17_N09_ZIG476-TEX5.ab1 852 0 852 ABI	Putative CCAAT-binding transcription factor [Oryza sativa (japonica cultivar-group)]	0
060711-17_N11_ZIG477-TEX5.ab1 844 0 844 ABI		0
060711-17_N17_ZIG480-TEX5.ab1 851 0 851 ABI	Diacylglycerol kinase 1 [Oryza sativa (japonica cultivar-group)]	1.11E-140
060711-17_N19_ZIG481-TEX5.ab1 865 0 865 ABI	Plastidic aspartate aminotransferase [Panicum miliaceum (Proso millet) (Broomcorn millet)]	0
060711-17_N21_ZIG482-TEX5.ab1 856 0 856 ABI		0
060711-17_N23_ZIG483-TEX5.ab1 855 0 855 ABI	40S ribosomal protein S7 [Secale cereale (Rye)]	0
060711-17_P03_ZIG485-TEX5.ab1 861 0 861 ABI	Expressed protein [Oryza sativa (japonica cultivar-group)]	0
060711-17_P05_ZIG486-TEX5.ab1 862 0 862 ABI		0
060711-17_P07_ZIG487-TEX5.ab1 852 0 852 ABI	Guanine nucleotide-exchange protein GEP2 [Oryza sativa (Rice)]	6.19E-113
060711-17_P09_ZIG488-TEX5.ab1 858 0 858 ABI	Cytosolic glyceraldehyde-3-phosphate dehydrogenase [Triticum aestivum (Wheat)]	0
060711-17_P11_ZIG489-TEX5.ab1 851 0 851 ABI		3.26E-05
060711-17_P13_ZIG490-TEX5.ab1 863 0 863 ABI	Hypothetical protein Atlg09150 [Arabidopsis thaliana (Mouse-ear cress)]	6.51E-137

060711-17_P17_ZIG492-TEX5.ab1 852 0	Hypothetical protein OJ1135_F06.7 [Oryza sativa (japonica cultivar-group)]	3.57E-06
060711-17_P19_ZIG493-TEX5.ab1 857 0	T6J4.13 protein [Arabidopsis thaliana (Mouse-ear cress)]	0
060711-17_P21_ZIG494-TEX5.ab1 852 0	Mitochondrial import inner membrane translocase subunit Tim9 [Oryza sativa (Rice)]	0
060711-17_P23_ZIG495-TEX5.ab1 856 0	Cytosolic heat shock protein 90 [Hordeum vulgare (Barley)]	0
zig100-tex5.ab1 1 407	Hypothetical protein OJ1513_F02.133 [Oryza sativa (japonica cultivar-group)]	7.68E-116
zig103-tex5.ab1 1 264	Hypothetical protein OJ1513_F02.133 [Oryza sativa (japonica cultivar-group)]	1.17E-89
zig104-tex5.ab1 1 361		1.16E-125
zig105-tex5.ab1 1 366	Putative annexin P35 [Oryza sativa (japonica cultivar-group)]	0
zig109-tex5.ab1 1 261	Ubiquitin family, putative [Oryza sativa (japonica cultivar-group)]	3.99E-130
zig110-tex5.ab1 1 429	Putative WD-40 repeat protein [Oryza sativa (japonica cultivar-group)]	1.72E-121
zig111-tex5.ab1 1 315	Hypothetical protein OJ1513_F02.133 [Oryza sativa (japonica cultivar-group)]	1.67E-113
zig112-tex5.ab1 1 429	Putative ribosomal RNA apurinic site specific lyase [Oryza sativa (japonica cultivar-group)]	4.78E-150
zig116-tex5.ab1 1 347	Hypothetical protein OJ1513_F02.133 [Oryza sativa (japonica cultivar-group)]	1.67E-113
zig117-tex5.ab1 1 224		1.74E-171
zig120-tex5.ab1 1 597	Cytosolic heat shock protein 90 [Hordeum vulgare (Barley)]	0
zig123-tex5.ab1 1 355	60S ribosomal protein L31 [Picea mariana (Black spruce)]	8.56E-70
zig127-tex5.ab1 1 307	Triosephosphate isomerase, cytosolic [Secale cereale (Rye)]	3.12E-179
zig128-tex5.ab1 1 532	Lactoylglutathione lyase [Oryza sativa (Rice)]	6.02E-134
zig12-tex5.ab1 1 494		5.48E-07
zig131-tex5.ab1 1 631	Hypothetical protein OJ1513_F02.133 [Oryza sativa (japonica cultivar-group)]	1.79E-113
zig132-tex5.ab1 1 296	Mitochondrial import inner membrane translocase subunit Tim9 [Oryza sativa (Rice)]	0
zig133-tex5.ab1 1 536	Glyceraldehyde-3-phosphate dehydrogenase, cytosolic [Hordeum vulgare (Barley)]	1.01E-141
zig137-tex5.ab1 1 490	60S ribosomal protein L10-3 [Oryza sativa (Rice)]	0
zig138-tex5.ab1 1 377		0
zig13-tex5.ab1 1 458	B1066G12.9 [Oryza sativa (japonica cultivar-group)]	0
zig142-tex5.ab1 1 244	putative esterase D [Oryza sativa (japonica cultivar-group)]	0
zig144-tex5.ab1 1 221	UDP-glucose:glycoprotein glucosyltransferase/ transferase, transferring glycosyl groups / transferase, transferring hexosyl groups [Arabidopsis thaliana]	8.18E-105
zig146-tex5.ab1 1 247	Disease-resistant-related protein [Oryza sativa (Rice)]	1.60E-93
zig148-tex5.ab1 1 519	Disease-resistant-related protein [Oryza sativa (Rice)]	1.20E-97

zig150-tex5.abl 1 491	Cytosolic heat shock protein 90 [Hordeum vulgare (Barley)]	0
zig151-tex5.abl 1 449	Putative polynucleotide adenylyltransferase [Oryza sativa (indica cultivar-group)]	0
zig154-tex5.abl 1 514	60S ribosomal protein L44-like [Arabidopsis thaliana (Mouse-ear cress)]	8.98E-134
zig156-tex5.abl 1 396		8.26E-153
zig158-tex5.abl 1 437	Ribosomal protein l34 [Triticum aestivum (Wheat)]	0
zig159-tex5.abl 1 431	Hypothetical protein OSJNBa0053E05.17 [Oryza sativa (japonica cultivar-group)]	0
zig160-tex5.abl 1 337	40 kDa PI 8.5 ABSCISSIC acid-induced protein [Oryza sativa (Rice)]	0
zig162-tex5.abl 1 268		5.13E-150
zig163-tex5.abl 1 268		1.22E-91
zig164-tex5.abl 1 491	Ribosomal protein L23 [Phytophthora infestans (Potato late blight fungus)]	2.87E-123
zig172-tex5.abl 1 386	Putative beta-1,3-glucanase [Sorghum bicolor (Sorghum) (Sorghum vulgare)]	2.91E-142
zig173-tex5.abl 1 352	ADP ribosylation GTPase-like protein [Oryza sativa (japonica cultivar-group)]	0
zig176-tex5.abl 1 448	Glyceraldehyde-3-phosphate dehydrogenase, cytosolic [Hordeum vulgare (Barley)]	1.18E-175
zig177-tex5.abl 1 622	Hypothetical protein P0413G02.21 [Oryza sativa (japonica cultivar-group)]	0
zig179-tex5.abl 1 439	MATE efflux family protein, putative [Oryza sativa (japonica cultivar-group)]	0
zig17-tex5.abl 1 334		1.57E-24
zig182-tex5.abl 1 295	Hypothetical protein P0036D10.5 [Oryza sativa (japonica cultivar-group)]	1.99E-162
zig186-tex5.abl 1 317	Putative ribosomal protein S29 [Oryza sativa (japonica cultivar-group)]	1.01E-141
zig187-tex5.abl 1 308		1.01E-98
zig188-tex5.abl 1 529	P-type ATPase [Hordeum vulgare (Barley)]	4.66E-150
zig192-tex5.abl 1 210	Putative male sterility 1 protein [Oryza sativa (japonica cultivar-group)]	1.30E-100
zig199-tex5.abl 1 495	Actin 1 [Avena nuda (Small naked oat)]	4.11E-63
zig19-tex5.abl 1 257	40S ribosomal protein S8 [Oryza sativa (Rice)]	0
zig200-tex5.abl 1 667	Enolase 2 [Zea mays (Maize)]	3.02E-173
zig203-tex5.abl 1 392	60S ribosomal protein L17-1 [Hordeum vulgare (Barley)]	0
zig205-tex5.abl 1 110	Peptidyl-prolyl cis-trans isomerase [Oryza sativa (japonica cultivar-group)]	0
zig206-tex5.abl 1 501	Putative actin depolymerizing factor [Oryza sativa (japonica cultivar-group)]	3.45E-34
zig208-tex5.abl 1 492	Expressed protein [Oryza sativa (japonica cultivar-group)]	0
zig209-tex5.abl 1 399		0
zig20-tex5.abl 1 370	Cytosolic heat shock protein 90 [Hordeum vulgare (Barley)]	0
zig213-tex5.abl 1 736	Hypothetical protein OSJNBb0027B08.11 [Oryza sativa (japonica cultivar-group)]	7.31E-54
zig214-tex5.abl 1 551		0
zig215-tex5.abl 1 548	T6D22.2 [Arabidopsis thaliana (Mouse-ear cress)]	0
zig220-tex5.abl 1 379	BRI1-KD interacting protein 108 [Oryza sativa (japonica cultivar-group)]	0
zig229-tex5.abl 1 347	SAC domain protein 7, putative [Oryza sativa (japonica cultivar-group)]	3.86E-154

	cultivar-group))]	
zig22-tex5.ab1 1 395		0
zig230-tex5.ab1 1 447	Cytosolic glyceraldehyde-3-phosphate dehydrogenase [Triticum aestivum (Wheat)]	3.79E-160
zig24-tex5.ab1 1 374	Hypothetical protein OJ1513_F02.133 [Oryza sativa (japonica cultivar-group)]	6.63E-110
zig253-tex5.ab1 1 561	Hypothetical protein OJ2056_H01.18 [Oryza sativa (japonica cultivar-group)]	0
zig254-tex5.ab1 1 105	E2F protein [Triticum sp]	0
zig255-tex5.ab1 1 440		1.24E-30
zig258-tex5.ab1 1 201		1.49E-175
zig259-tex5.ab1 1 392		1.24E-81
zig262-tex5.ab1 1 322	60s ribosomal protein L21 [Triticum aestivum (Wheat)]	1.11E-122
zig265-tex5.ab1 1 578	60s ribosomal protein L21 [Triticum aestivum (Wheat)]	5.47E-159
zig267-tex5.ab1 1 456	Putative speckle-type POZ protein [Oryza sativa (japonica cultivar-group)]	5.89E-06
zig268-tex5.ab1 1 278	Hypothetical protein OJ1513_F02.133 [Oryza sativa (japonica cultivar-group)]	2.99E-118
zig26-tex5.ab1 1 262	Proteasome subunit beta type 3-A [Arabidopsis thaliana (Mouse-ear cress)]	0
zig271-tex5.ab1 1 480	Putative valyl tRNA synthetase [Oryza sativa (japonica cultivar-group)]	0
zig272-tex5.ab1 1 257	Hypothetical protein OJ1513_F02.133 [Oryza sativa (japonica cultivar-group)]	1.87E-113
zig273-tex5.ab1 1 336		2.27E-71
zig274-tex5.ab1 1 231	Expressed protein [Oryza sativa (japonica cultivar-group)]	1.62E-168
zig276-tex5.ab1 1 621	Hypothetical protein [Arabidopsis thaliana (Mouse-ear cress)]	3.05E-104
zig278-tex5.ab1 1 395	T6D22.2 [Arabidopsis thaliana (Mouse-ear cress)]	0
zig27-tex5.ab1 1 443	Lactoylglutathione lyase [Oryza sativa (Rice)]	1.32E-131
zig280-tex5.ab1 1 465	Dehydrin [Hordeum vulgare (Barley)]	2.58E-155
zig285-tex5.ab1 1 267		0
zig286-tex5.ab1 1 666	RuBisCO large subunit-binding protein subunit beta, chloroplast [Secale cereale (Rye)]	1.58E-112
zig288-tex5.ab1 1 494	Hypothetical protein T9A21.50 [Arabidopsis thaliana (Mouse-ear cress)]	0
zig291-tex5.ab1 1 242	Putative eukaryotic translation initiation factor 3, subunit 1 alpha, 35kDa [Oryza sativa (japonica cultivar-group)]	3.19E-111
zig292-tex5.ab1 1 353	Putative RNA-binding protein 10 [Oryza sativa (japonica cultivar-group)]	4.31E-91
zig294-tex5.ab1 1 261	Disease-resistant-related protein [Oryza sativa (Rice)]	2.87E-173
zig296-tex5.ab1 1 332		1.10E-122
zig29-tex5.ab1 1 341	Putative mitogen-activated protein kinase kinase [Oryza sativa (Rice)]	0
zig300-tex5.ab1 1 553	T6D22.2 [Arabidopsis thaliana (Mouse-ear cress)]	4.49E-78
zig302-tex5.ab1 1 192	40S ribosomal protein S11-1 [Arabidopsis thaliana (Mouse-ear cress)]	0
zig303-tex5.ab1 1 247	40S ribosomal protein S11-1 [Arabidopsis thaliana (Mouse-ear cress)]	0
zig306-tex5.ab1 1 307	Hypothetical protein At5g10750 [Arabidopsis thaliana (Mouse-ear cress)]	6.85E-53
zig307-tex5.ab1 1 464	Hypothetical protein F24M12.290 [Arabidopsis thaliana (Mouse-ear cress)]	4.48E-147

zig309-tex5.abl 1 455	Acidic ribosomal protein [Triticum aestivum (Wheat)]	0
zig30-tex5.abl 1 312		5.36E-08
zig311-tex5.abl 1 368	Proteasome inhibitor-like protein [Oryza sativa (japonica cultivar-group)]	0
zig312-tex5.abl 1 199	40S ribosomal protein S8 [Oryza sativa (Rice)]	0
zig313-tex5.abl 1 545	Cytosolic glyceraldehyde-3-phosphate dehydrogenase [Triticum aestivum (Wheat)]	8.44E-52
zig316-tex5.abl 1 278		0
zig320-tex5.abl 1 322	Putative ATP-dependent DNA helicase recQ [Oryza sativa (japonica cultivar-group)]	3.30E-21
zig323-tex5.abl 1 527	Hypothetical protein OJ1513_F02.133 [Oryza sativa (japonica cultivar-group)]	2.80E-118
zig325-tex5.abl 1 428		0
zig326-tex5.abl 1 483	Hypothetical protein OJ1513_F02.133 [Oryza sativa (japonica cultivar-group)]	2.80E-118
zig327-tex5.abl 1 397	Hypothetical protein OJ1513_F02.133 [Oryza sativa (japonica cultivar-group)]	7.71E-116
zig331-tex5.abl 1 345	Hypothetical protein OJ1513_F02.133 [Oryza sativa (japonica cultivar-group)]	2.87E-118
zig332-tex5.abl 1 494	Cytosolic glyceraldehyde-3-phosphate dehydrogenase [Triticum aestivum (Wheat)]	2.80E-173
zig333-tex5.abl 1 242	Cytosolic heat shock protein 90 [Hordeum vulgare (Barley)]	0
zig335-tex5.abl 1 280		1.02E-122
zig338-tex5.abl 1 305	Malate dehydrogenase [Oryza sativa (Rice)]	1.33E-131
zig339-tex5.abl 1 404		3.17E-120
zig33-tex5.abl 1 438	Hypothetical protein P0459B01.12 [Oryza sativa (japonica cultivar-group)]	0
zig342-tex5.abl 1 331	P0470A12.7 [Oryza sativa (japonica cultivar-group)]	0
zig346-tex5.abl 1 316	Lactoylglutathione lyase [Oryza sativa (Rice)]	9.13E-161
zig347-tex5.abl 1 436	Lactoylglutathione lyase [Oryza sativa (Rice)]	4.13E-138
zig350-tex5.abl 1 268	Hypothetical protein OJ1513_F02.133 [Oryza sativa (japonica cultivar-group)]	6.96E-116
zig352-tex5.abl 1 210	Putative zinc finger transcription factor ZFP33 [Oryza sativa (japonica cultivar-group)]	4.48E-122
zig353-tex5.abl 1 302		3.69E-91
zig359-tex5.abl 1 330		1.29E-122
zig360-tex5.abl 1 306	Ribosomal protein L35A [Zea mays (Maize)]	3.10E-148
zig363-tex5.abl 1 366	Lactoylglutathione lyase [Oryza sativa (Rice)]	1.36E-125
zig364-tex5.abl 1 339	OSJNBa0006A01.14 protein [Oryza sativa (Rice)]	0
zig367-tex5.abl 1 361	Putative 60S ribosomal protein L12 [Oryza sativa (japonica cultivar-group)]	6.71E-171
zig368-tex5.abl 1 230	Cytosolic glyceraldehyde-3-phosphate dehydrogenase [Triticum aestivum (Wheat)]	4.95E-178
zig372-tex5.abl 1 508	Putative RNA-binding protein 10 [Oryza sativa (japonica cultivar-group)]	1.55E-87
zig374-tex5.abl 1 388	Ribosomal protein L3 [Triticum aestivum (Wheat)]	0
zig375-tex5.abl 1 288	Putative nitrilase-associated protein [Arabidopsis thaliana (Mouse-ear cress)]	0
zig377-tex5.abl 1 326		9.46E-139
zig380-tex5.abl 1 374	Disease-resistant-related protein [Oryza sativa (Rice)]	2.36E-114
zig381-tex5.abl 1 615	Cysteine proteinase inhibitor [Triticum aestivum (Wheat)]	0
zig38-tex5.abl 1 248	Putative ribosomal protein L34 [Oryza sativa (japonica cultivar-]	1.46E-162

	group)]	
zig40-tex5.ab1 1 560	Putative heat shock factor binding protein [Oryza sativa (japonica cultivar-group)]	0
zig42-tex5.ab1 1 414	Putative S-locus protein 4 [Oryza sativa (japonica cultivar-group)]	3.21E-170
zig44-tex5.ab1 1 285	Ribosomal protein l34 [Triticum aestivum (Wheat)]	0
zig49-tex5.ab1 1 319	ATP synthase D chain, mitochondrial [Arabidopsis thaliana (Mouse-ear cress)]	0
zig50-tex5.ab1 1 299	Hypothetical protein OJ1513_F02.133 [Oryza sativa (japonica cultivar-group)]	1.19E-114
zig56-tex5.ab1 1 466	Hypothetical protein OJ1513_F02.133 [Oryza sativa (japonica cultivar-group)]	1.16E-117
zig57-tex5.ab1 1 333	Hypothetical protein OJ1014_C08.10 [Oryza sativa (japonica cultivar-group)]	0
zig63-tex5.ab1 1 368		3.91E-166
zig65-tex5.ab1 1 499	Putative zinc finger transcription factor ZFP33 [Oryza sativa (japonica cultivar-group)]	4.87E-175
zig67-tex5.ab1 1 381	Hypothetical protein At3g11800 [Arabidopsis thaliana (Mouse-ear cress)]	0
zig71-tex5.ab1 1 404	Hypothetical protein [Arabidopsis thaliana (Mouse-ear cress)]	2.44E-164
zig73-tex5.ab1 1 288	Putative immunophilin [Hordeum vulgare (Barley)]	0
zig74-tex5.ab1 1 404	Putative puromycin-sensitive aminopeptidase [Oryza sativa (japonica cultivar-group)]	7.28E-118
zig77-tex5.ab1 1 487	Putative organic solute transporter [Oryza sativa (Rice)]	0
zig7-tex5.ab1 1 336	Ribosomal protein L15 [Oryza sativa (japonica cultivar-group)]	0
zig83-tex5.ab1 1 324	Putative salt tolerance protein 5 [Oryza sativa (japonica cultivar-group)]	0
zig85-tex5.ab1 1 323	26S proteasome regulatory particle non-ATPase subunit12 [Oryza sativa (japonica cultivar-group)]	2.92E-145
zig88-tex5.ab1 1 485	26S proteasome regulatory particle non-ATPase subunit12 [Oryza sativa (japonica cultivar-group)]	1.11E-104
zig90-tex5.ab1 1 452	Hypothetical protein OJ1513_F02.133 [Oryza sativa (japonica cultivar-group)]	3.18E-118
zig92-tex5.ab1 1 467	Eukaryotic translation initiation factor 5A1 [Triticum aestivum (Wheat)]	0
zig95-tex5.ab1 1 561	Hypothetical protein OJ1513_F02.133 [Oryza sativa (japonica cultivar-group)]	3.07E-118
zig98-tex5.ab1 1 467	Putative copper chaperone [Oryza sativa (japonica cultivar-group)]	0
zig99-tex5.ab1 1 455	Hypothetical protein OJ1513_F02.133 [Oryza sativa (japonica cultivar-group)]	3.07E-118

12. 3. Maghoz köthető EST könyvtárak listája

könyvtár leírás	szövet	fejlődési állapot	fajta
DNA amplification by PCR	endosperm	developing endosperm	
wheat developing seed cDNA library		embryo	
pcTA5		immature endosperm	
pcTA9		immature endosperm	

lambda gt10	endosperm		
Mike Holdsworth, Plant Science, IACR, Bristol		embryo	
developing endosperm cDNA cloned in pUC12	endosperm	developing endosperm	
lambda Charon 32	endosperm		
Wheat Spikelet	Spikelet		
SSH-cDNA Library of young spike of YS type	spike	dinucleate stage	
differential displayed sequences in spikelets	spikelets	after the miosis of Pollen mother cell	
DDRT-RCHW	Spikelet		
spikelets of YS thermo-sensitive male sterile	spikelets	after the miosis of Pollen mother cell	
wip1c	pistils		
wdk2c	kernel		
wdk1c	kernel		
wialc	anthers		
wem1c	embryo		
wde2f	endosperm		
wde1f	endosperm		
wdk3c	kernel		
wpa1c	anthers		
wheat seed cDNA library	seed		
PSR	embryo		
AAFC_CRC Fusarium graminearum inoculated wheat	spike	anthesis	93FHB37
Ta02_AAFC_ECORC_Fusarium_graminearum_inoculate	heads	anthesis	93FHB37
DDRT PCR product	embryo	35 days after flowering	AUS1408
Triticum aestivum floret differentiation to	young ear	floret differentiation to anther separation	BAU3338
genomic DNA		mature seed	Begra
Lambda ZAP (Stratagene)	embryonic axis	seed, dormant	Brevor
abscisic acid-treated embryo cDNA library		mature embryo	Brevor
Brevor wheat dormant embryo cDNA library	Seed embryo	Mature seed	Brevor
Wheat dormant embryo cDNA library	Seed embryo	Mature seed	Brevor
Wheat ABA-treated embryo cDNA library	Seed embryo	Mature dormant seeds	Brevor (soft, white, winter, common wheat)
TAcDNA0025 Lot#1.1.L, TA059E1X, obtained from		developing grain	Butte 86
Wheat developing grains cDNA library	whole grains	3-44 days post anthesis seed	Butte 86

cDNA	seed	23 DAF	Capitole
lambda /R20		embryonic	Cheyenne
lambda/wheat genomic DNA	endosperm	seed	Cheyenne
lambda-Sep6-CNN		embryo	Cheyenne
Cheyenne wheat endosperm cDNA library	Cheyenne endosperm	5 -30 days post anthesis seed	Cheyenne
Wheat subtracted endosperm cDNA library	Endosperm	5 -30 days post anthesis seed	Cheyenne
Wheat endosperm cDNA library	Endosperm	5 to 30 days post anthesis seed	Cheyenne
ITEC WHE Wheat Endosperm Library	endosperm	5-30 days post anthesis	Cheyenne
DNA amplification by PCR library	endosperm	developing endosperm	Chinese Spring
lambda-Charen 34		embryo	Chinese Spring
lambda ZAPII	endosperm	endosperm, immature	Chinese Spring
pBR322		de-embryonated seeds	Chinese Spring
Y. Ogihara unpublished library	spike at flowering date	Feekes' scale 10.5.1	Chinese Spring
plasmid wheat spike cDNA library		Flowering stage	Chinese Spring
pST3, N terminal by PCR	Developing endosperm	15 DAP	Chinese Spring
pST5, N-terminal by PCR	developing endosperm	15 D.A.P.	Chinese Spring
cDNA library in lambda ZapII		developing grain	Chinese Spring
cDNA-AFLP screen, Chinese Spring	pistil		Chinese Spring
Y.Ogihara unpublished cDNA library Wh_FL	spikelet	early flowering	Chinese Spring
Y.Ogihara unpublished cDNA library Wh_DPA20	seed DPA20		Chinese Spring
Y.Ogihara unpublished cDNA library Wh_OKCS	seed DPA5		Chinese Spring
Y.Ogihara unpublished cDNA library Wh_PCDAM	anther at meiosis		Chinese Spring
Y. Ogihara unpublished cDNA library, Wh_e	seed DPA10	Feekes' scale 11.2	Chinese Spring
Y. Ogihara unpublished cDNA library, Wh_h	spike at heading date	Feekes' scale 10.5	Chinese Spring
Y. Ogihara unpublished cDNA library, Wh_f	spike at flowering date	Feekes' scale 10.5.1	Chinese Spring
Y. Ogihara unpublished cDNA library, Wh	spike at meiosis	Feekes' scale 9	Chinese Spring
Chinese Spring Differential display product	immature spike	Feekes growth stage 9-10	Chinese Spring

Chinese Spring wheat immature spikelet cDNA		3-10 mm	Chinese Spring
Differential display fragments A	Anthers	preanthesis	Chinese Spring
CS wheat heat stressed spike cDNA library	Whole spike	Spikes at 5, 10, 15 and 20 days after	Chinese Spring
CS wheat pre-anthesis spike cDNA library	Spike before anthesis	Adult plant	Chinese Spring
CS wheat 5-15 DAP spike cDNA library	Spike	Adult plant	Chinese Spring
CS wheat 20-45 DAP spike cDNA library	Spike and seed	Adult plant	Chinese Spring
waw1c	anthers		Chinese Spring
Wheat meiotic anther cDNA library	Anther	Meiotic stages pre-meiosis-metaphase I	Chinese Spring
Y. Ogihara unpublished cDNA library, Wh_yf	spikelet at early flowering	Feekes' scale 6	Chinese Spring
Y. Ogihara unpublished cDNA library, Wh_yd	spikelet at late flowering	Feekes' scale 6	Chinese Spring
Y. Ogihara unpublished cDNA library, Wh_SL	seed DPA30	Feekes' scale 11.3	Chinese Spring
Y. Ogihara unpublished cDNA library, Wh_oh	pistil at heading date	Feekes' scale 10.5	Chinese Spring
Wheat heat stressed spike cDNA library	Whole spike	Spikes at 5, 10, 15 and 20 days after	Chinese Spring
Wheat pre-anthesis spike cDNA library	Spike before anthesis	Adult plant	Chinese Spring
Wheat 5-15 DAP spike cDNA library	Spike	Adult plant	Chinese Spring
Wheat 20-45 DAP spike cDNA library	Spike and seed	Adult plant	Chinese Spring
Differential display fragments O	Ovaries	preanthesis	Chinese Spring
Ta10c_AAFC_ECORC_Fusarium_graminearum_inculate	spikelets	4 days after anthesis	Chinese Spring addition line contain. chom 7E
Ta10b_AAFC_ECORC_Fusarium_graminearum_inculate	spikelets	4 days after anthesis	Chinese Spring addition line contain. chom 7E
Ta10_AAFC_ECORC_Fusarium_graminearum_inculated	spikelets	4 days after anthesis	Chinese Spring addition line contain. chom 7E

Ta09_AAFC_ECORC_Fusarium_graminearum_inoculated	spikelets	anthesis	Chinese Spring addition line contain. chom 7E
Y.Ogihara unpublished cDNA library	anther at meiosis		DMT4B(CS)
Y.Ogihara unpublished cDNA library Wh_MS	seed DPA5		DT3DL(CS)
Y.Ogihara unpublished cDNA library Wh_GCPCDAM	anther at meiosis		DT4B(CS)
Wheat-Fusarium graminearum filtrate	Embryo	immature grain, 10 hr post inoculation	E2-24-T
Ta03_AAFC_ECORC_Fusarium_graminearum_inoculate	heads	anthesis	FHB148
EC		unfertilized	Florida
q:444		24 hours after fertilization	Florida
q:242	egg cell	pre-fertilisation	Florida
q:343	ovules	pre-fertilization	Florida
q:141	ovules	2 days post fertilization	Florida
Ta01_AAFC_ECORC_Fusarium_graminearum_inoculate	heads	anthesis	Frontana
TaE05	developing seeds	5 days after anthesis	Glenlea
TaE15	developing seeds	15 days after anthesis	Glenlea
wkm2c	kernel		hard red spring
wkm1c	kernel		hard red spring
Hartog wheat EST endosperm library	endosperm	developing endosperm tissue 6, 8, 10 dpa (days	Hartog
Wheat EST endosperm library	endosperm	developing endosperm tissue 6, 8, 10 dpa (days	Hartog
Ta04_AAFC_ECORC_Fusarium_graminearum_inoculate	heads	anthesis	Harus
lambda NM1130/wheat genomic	endosperm	seed	Hope
Y.Ogihara unpublished cDNA library Wh_EMC	dormant seed with cold treatment after water		kitaKEI1354
Y.Ogihara unpublished cDNA library Wh_EM	dormant seed with water absorption		kitaKEI1354
Y.Ogihara unpublished cDNA library Wh_EMI	dormant seed with water absorption after		kitaKEI1354

LambdaZapScut	scutellum	germinating grain 3 days at 25 degrees C	Maris Huntsman
e:411	maternal tissue (minus endosperm and	6 days post anthesis	Mercia
e:512	maternal tissue (minus endosperm and	8 days post anthesis	Mercia
e:29	endosperm	8 days post anthesis	Mercia
j:122	embryo	14 days post anthesis	Mercia
j:223	embryo	21 days post anthesis	Mercia
h:116	endosperm	14 days post anthesis	Mercia
e:310	endosperm	10 days post anthesis	Mercia
d:15	embryo (micropylar tip)	6 days post anthesis	Mercia
d:37	embryo (micropylar tip)	10 days post anthesis	Mercia
d:26	embryo (micropylar tip)	8 days post anthesis	Mercia
a:11	carpel tissue	24 hours post anthesis	Mercia
a:22	carpel tissue	3-4 days post anthesis	Mercia
n:129	embryo	1 day post germination	Mercia
j:324	embryo	28 days post anthesis	Mercia
l:226	embryo	40 days post anthesis	Mercia
N:130	embryo	2 days post germination	Mercia
O:232	aleurone	2 days post germination	Mercia
l:125	embryo	30 days post anthesis	Mercia
lambda ZAP wheat young spike cDNA	young spike		Norin 26
lambdaZAP wheat spike cDNA library		Floret differentiation stage	Norin 26
ITEC MUG Wheat Spikelet Library	young spikelets	Feekes' scale 6-7	Norin 26
ITEC PSR Wheat Pericarp/Testa Library	pericarp/testa	3:1 mix 10:20 days post anthesis	Novosibirsk aya 67
ITEC WWS Wheat Scutellum Library	scutellum callus		Novosibirsk aya 67
OV	ovary		recital
GR45	grain (45 degrees per day after		recital
G750	grain (750 degrees per day after		recital
G608	grain (608 degrees per day after		recital

G550	grain (550 degrees per day after		recital
G468	grain (468 degrees per day after		recital
G356	grain (356 degrees per day after		recital
G174	grain (174 degrees per day after		recital
G118	grain (118 degrees per day after		recital
G100R	grain (45 degrees per day after		recital
G100D	grain (118 degrees per day after		recital
Ta08_AAFC_ECORC_Fusarium_graminearum_inculated	spikelets	2 days after anthesis	Roblin
Ta07b_AAFC_ECORC_Fusarium_graminearum_inculate	spikelets	4 days after anthesis	Roblin
Ta07_AAFC_ECORC_Fusarium_graminearum_inculated	spikelets	4 days after anthesis	Roblin
total soft falcon	endosperm	mid	soft falcon
ITEC PSR Wheat Endosperm Library	endosperm	1:1 mix 10:20 days post anthesis	Soleil
wdk4c	kernel		Spring wheat
wdk9n1	kernel		Spring wheat
wdk9n	kernel		Spring wheat
wkm2n	kernel		Spring wheat
wdk5c	kernel		Spring wheat
Wheat Fusarium graminearum infected spike cDNA	Spike	Adult plant	Sumai3
Sumai3 Wheat Fusarium graminearum infected	Spike	Adult plant	Sumai3
KSU wheat Fusarium graminearum infected spike	Spike	Adult plant	Sumai3
ITEC SCU Wheat Endosperm Library	endosperm		Wyuna
ITEC CSB Wheat Endosperm Library	endosperm	8-12 days post anthesis	Wyuna